

WICC 2016

XVIII Workshop de Investigadores en Ciencias de la Computación

14 y 15 de abril · Concordia · Entre Ríos

AREAS DE INTERES

- Agentes y Sistemas Inteligentes
- Arquitecturas, Redes y Sistemas Operativos
- Bases de Datos y Minería de Datos
- Computación Gráfica, Imágenes y Visualización
- Ingeniería de Software
- Innovación en Educación en Informática
- Innovación en Sistemas de Software
- Procesamiento Distribuido y Paralelo
- Procesamiento de Señales y Sistemas de Tiempo Real
- Seguridad Informática
- Tecnología Informática Aplicada en Educación
- Tesis Doctorales

ISBN 978-950-698-377-2



Red de Universidades
con Carreras en Informática



Facultad de Ciencias
de la **Administración**

Autoridades

Universidad Nacional de Entre Ríos

Rector: Ing. Jorge Gerard

Facultad de Ciencias de la Administración

Decano: Cr. Hipólito Buenaventura FINK

Red de Universidades Nacionales con Carreras de Informática (Red UNCI)

Coordinador Titular

Guillermo Feierherd (UNTDF)

Coordinador Alterno

Patricia Pesado (UNLP)

Junta Directiva

- Fabiana Piccoli (UNSL) 2014-2016
- Marcelo Estayno (UNLZ) 2014-2016
- Gladys Dapozo (UNNE) 2014-2016
- Hugo Padovani (U Morón) 2014-2016
- Horacio Kuna (UN Misiones) 2015-2017
- Jorge Finocchietto (U CAECE) 2015-2017
- Osvaldo Sposito (UN La Matanza) 2015-2017
- Claudia Russo (UNNOBA) 2015-2017

Secretarías

- Secretaría Administrativa: Javier Balladini (UNCOMA)
- Secretaría Académica: Marta Lasso (UNPA)
- Secretaría de Congresos, Publicaciones y Difusión: Armando De Giusti (UNLP)
- Secretaría de Asuntos Reglamentarios: Marcelo Devincenzi (UAI)

Comité Académico WICC 2016

Universidades	Representante
UBA – Cs. Exactas	Fernández Slezak Diego
UBA – Ingeniería	Echeverría, Adriana
UN La Plata	Pesado, Patricia
UN Sur	Rueda Sonia
UN San Luis	Piccoli, Fabiana
UNCPBA	Aciti Claudio
UN Comahue	Balladini, Javier
UN La Matanza	Sposito, Osvaldo
UN La Pampa	Alfonso, Hugo
UN Lomas de Zamora	Estayno, Marcelo
UN Tierra del Fuego	Feierherd, Guillermo
UN Salta	Gil, Gustavo
UN Patagonia Austral	Márquez, María Eugenia
UN San Juan	Otazú, Alejandra
UADER	Aranguren, Silvia

UN Patagonia SJB	Buckle, Carlos
UN Entre Ríos	Tugnarelli, Mónica
UN Nordeste	Dapozo, Gladys
UN Rosario	Kantor Raul
UN Misiones	Kuna, Horacio
UNNOBA	Russo, Claudia
UN Chilecito	Carmona, Fernanda
UN Lanús	García Martínez, Ramón
UN Santiago del Estero	Duran Elena
Esc. Sup. Ejército	Castro Lechstaler Antonio
UN Litoral	Loyarte, Horacio
UN RioIV	Arroyo, Marcelo
UN Córdoba	Fridlender, Daniel
UN Jujuy	Herrera Cognetta, Analía
UN Rio Negro	Vivas, Luis
UN Villa María	Prato, Laura
UN Lujan	Scucimarri, Jorge
UN Catamarca	Barrera, María Alejandra
UN La Rioja	Campazzo Eduardo
UNTres de Febrero	Oliveros, Alejandro
UN Tucumán	Luccioni, Griselda María
UNAJ	Morales, Martín
UN Chaco Austral	Zachman Patricia
UN del Oeste	Foti, Antonio
UN de Cuyo	Forradelas, Raymundo
UNdeMardelPlata	Doumecq, Julio Cesar
UNdeQuilmes	Díaz-Caro, Alejandro
UMorón	Padovani Hugo
UAI	De Vincenzi, Marcelo
UBelgrano	Guerci, Alberto
Ukennedy	Panizzi, Marisa
UAdventistadelPlata	Bournissen Juan
UCAECE	Finocchieto, Jorge
UPalermo	Alvarez Adriana
UCARosario	Grieco, Sebastián
USalvador	Zanitti, Marcelo
UAconcagua	Giménez, Rosa
UGastónDachary	Beyersdorf, Carlos
UCEMA	Guglianone, Ariadna
UAustral	Cassol, Ignacio
UAtlántida Argentina	Rathmann, Liliana
UCALaPlata	Bertone, Rodolfo
ITBA	Gomez, Silvia
UChampagnat	Pincioli, Fernando

Coordinadores de Área WICC 2016

Agentes y Sistemas Inteligentes

Daniel Pandolfi (UNPA)
Marcelo Falappa (UNS)
Marcelo Errecalde (UNSL)

Arquitectura, Redes y Sistemas Operativos

Luis Marrone (UNLP)
Hugo Padovani (UMorón)
Daniel Arias Figueroa (UN Salta)

Computación Gráfica, Imágenes y Visualización

Martín Larrea (UNS)
María J. Abásolo (UNLP-UNCPBA)
Roberto Guerrero (UNSL)

Ingeniería de Software

Pablo Fillotrani (UNSur)
Pablo Thomas (UNLP)
Fernanda Carmona (UNChilecito)

Innovación en Sistemas de Software

Osvaldo Sposito (UNLaM)
Gladys Dapozo (UNNE)
Marcelo Estayno (UNLZ)
Hugo Ramón (UNNOBA)

Tecnología Informática aplicada en Educación

Zulma Cataldi (UBA-UTN)
Alejandra Zangara (UNLP)
Mónica Tugnarelli (UNER)
Gustavo Gil (UNSalta)

Procesamiento Distribuido y Paralelo

Marcelo Naiouf (UNLP)
Marcela Printista (UNSL)
Javier Ballardini (UNCOMA)

Procesamiento de señales y Sistemas de Tiempo Real

Oscar Bría (INVAP)
Fernando Tinetti (UNLP)
Nelson Acosta (UNCPBA)

Bases de Datos y Minería de Datos

Laura Lanzarini (UNLP)
Claudia Deco (UNR)
Verónica Gil Costa (UNSL)

Innovación en Educación Informática

Jorge Finocchietto (UCAECE)
Claudia Russo (UNNOBA)
Elena Durán (UNSE)

Seguridad Informática

Paula Venosa (UNLP)
Javier Echaiz (UNSur)
Antonio Castro Lechtaller (IESE)

Jurado Tesis Doctorado

Marcela Printista (UNSL)
Laura De Giusti (UNLP)
Silvia Castro (UNS)
Alejandra Cechich (UNCOMA)
Horacio Kuna (UNaM)
Orlando Micolini (UNC)

AGENTES Y SISTEMAS INTELIGENTES

<i>Utilización de técnicas subsimbólicas de la inteligencia artificial para la generación de energías limpias</i>	14
<i>Operadores de Cambio Múltiples en Bases de Creencias Bajo Lógica Horn</i>	19
<i>Cambio de Creencias Múltiples en Sistemas Argumentativos Aplicado para Programación Lógica Rebatible</i>	24
<i>Técnicas de razonamiento bajo incertidumbre aplicadas a la estimación de los destinos de viajes de ruta simple</i>	29
<i>Técnicas de Inteligencia Artificial aplicadas a problemas científico-tecnológicos</i>	33
<i>Algoritmos Evolutivos híbridos para el diseño y operación eficiente de una red de distribución de agua potable</i>	38
<i>Algoritmo CHC aplicado a la optimización de coberturas de señales de radio frecuencia en comunicaciones inalámbricas en locaciones petroleras</i>	43
<i>Desarrollo de herramientas para la operabilidad de procesos productivos</i>	48
<i>Benchmarks para Problemas de Scheduling de Máquinas Paralelas Idénticas con Algoritmos Inteligentes</i>	53
<i>Toma de decisiones y aprendizaje en agentes artificiales inteligentes</i>	56
<i>Introducir el Manejo de Múltiples Criterios de Comparación de Argumentos en Sistemas Argumentativos</i>	61
<i>Definición de la infraestructura para procesos masivos de argumentación mediante aplicación de revisión de creencias y argumentación sobre ontologías Datalog+/-</i>	66
<i>Razonamiento basado en la tolerancia a la inconsistencia</i>	71
<i>Propuesta de una Arquitectura Argumentativa Flexible en DeLP</i>	76
<i>Arquitectura General de un Sistema Estructurador de Argumentos</i>	81
<i>Entornos Argumentativos Dinámicos</i>	86
<i>Herramientas para representación y razonamiento con preferencias bajo incertidumbre</i>	91
<i>Algoritmos Genéticos Distribuidos: Heterogeneidad y Migración</i>	96
<i>Tratamiento de Metas en Planificación Continua</i>	99
<i>Servicios Web combinados con Agentes Inteligentes</i>	103
<i>Representación y razonamiento con bases de conocimiento probabilísticas</i>	107
<i>Argumentación probabilística y revisión de creencias con aplicaciones a ciberseguridad</i>	112
<i>Modelos de Análisis de Información para la Toma de Decisiones Estratégicas del Sector Tealero</i>	117
<i>Optimización de rutas en el transporte de personas para la pequeña y mediana empresa</i>	122
<i>Búsqueda local iterada para resolver problemas de planificación</i>	127

<i>Tratamiento Masivo de Datos Utilizando Técnicas de Machine Learning</i>	131
<i>Aplicación de Minería de Datos Espacial en el Área de Salud en la zona de influencia de la UNNOBA</i>	135
ARQUITECTURA, SISTEMAS OPERATIVOS Y REDES	
<i>Análisis Comparativo de métricas de QoS de Redes Wi-Fi 802.11e</i>	139
<i>Evaluación del Enrutamiento y Acceso al Medio en Redes Mesh Auto-Configurables</i>	144
<i>Eficiencia Energética en Dispositivos Móviles para Facilitar su Uso en Zonas Rurales Aisladas</i>	149
<i>Redes de Sensores Inalámbricos en Entornos Industriales para monitorear Condiciones de Higiene y Seguridad</i>	154
<i>Arquitecturas adaptadas para integrar computación móvil y computación en la nube</i>	159
<i>Análisis de Eficiencia en Arquitecturas Multiprocesador para Aplicaciones de Transmisión y Procesamiento de Datos</i>	163
<i>Plataforma Middleware para la Gestión Datos de WSN de Manera Interoperable</i>	169
<i>Desarrollo de la comunicación Peer to Peer para Sistemas de Tiempo Real basado en IPV6</i>	174
<i>Explorando posibles mejoras de protocolo TCP en redes móviles</i>	177
<i>La conectividad en las zonas rurales</i>	182
<i>Herramientas de Software de Simulación para Redes de Comunicaciones</i>	189
<i>Evaluación de performance en Redes Definidas por Software para Centro de Datos – Hipótesis de Trabajo</i>	193
<i>Controlador DNP3 para la CIAA</i>	198
<i>RED SIPIA-LP Estudio de mecanismos de bajo consumo energético aplicados a Red de Sensores Inalámbricos en el ámbito de Agricultura de Precisión.</i>	202
<i>GW-CIAA-IoT: Gateway con CIAA para red inalámbrica de IoT</i>	206
<i>Prolongando Vida Útil de Redes de Sensores. Modificación Protocolo AODV</i>	211
BASES DE DATOS Y MINERÍA DE DATOS	
<i>Clasificación por Enterotipos y Grupos Ortólogos del Microbioma Humano con Métodos No Supervisados</i>	218
<i>Extracción de Conocimiento en el Cursado del Ciclo Común de Articulación de Carreras de Ingeniería</i>	223
<i>Grandes Datos y Algoritmos Eficientes para Búsquedas de Escala Web</i>	227
<i>Análisis, interpretación y toma de decisiones estratégicas en la Ciencia de Datos</i>	233
<i>Extracción y clasificación automatizada para la Unidad de Vigilancia Tecnológica e Inteligencia Competitiva de la Patagonia</i>	238

<i>Arquitectura Genérica para el Almacenamiento de Datos Biométricos</i>	243
<i>Técnicas de Clasificación aplicadas al rendimiento académico</i>	248
<i>Evolución del Modelo Multidimensional en un DataWarehouse para pacientes diabéticos</i>	253
<i>Metodología para Evaluación de Impacto de Migración entre Versiones de Bases de Datos NoSQL</i>	258
<i>Personalización de la Educación a través de la creación de Perfiles dinámicos de los alumnos</i>	263
<i>Minería de Textos y de la Web</i>	268
<i>Framework para Data Mining Educativo: Formalización y Aplicaciones</i>	273
<i>Framework basado en Ontologías para la Recuperación de Imágenes Médicas</i>	278
<i>Búsquedas en Grandes Volúmenes de Datos</i>	283
<i>Minería de Datos Aplicada a Datos Masivos</i>	288
<i>NoSql en sistemas distribuidos sobre Cluster Hadoop</i>	293
<i>Estudio de Integración de Métodos de Descubrimiento de Conocimiento en Web</i>	298
<i>Aprendizaje Automático aplicado a Reconocimiento de Patrones en Video y Minería de Datos</i>	303
<i>Aplicaciones del aprendizaje automático en psicología y educación</i>	308
<i>Aportes al Modelo de Bases de Datos Métricas</i>	313
<i>Geometría Computacional y Bases de Datos</i>	318
<i>Plataforma Web para acceder a diferentes Sistemas de Gestión de Bases de Datos Relacionales</i>	323
<i>Big Data y su impacto sobre las comunidades</i>	328
<i>Sentiment Analysis para la clasificación de noticias financieras en los Mercados Argentinos. Un modelo híbrido de POST enriquecido semánticamente</i>	333
<i>Comportamiento de Bases de Datos No Relacionales en Entornos Distribuidos</i>	338
<i>Indexación y Búsqueda sobre Datos no Estructurados</i>	343
<i>Confluencia de Áreas de Conocimiento en un Laboratorio de Sistemas Inteligentes</i>	348
<i>Evaluación de técnicas de Data Mining para la obtención de perfiles de ingresantes a la UNNOBA</i>	353
COMPUTACIÓN GRÁFICA, IMÁGENES Y VISUALIZACIÓN	
<i>Algoritmos para el tratamiento de imágenes de campos de luz</i>	356
<i>Generación de Características y Reconocimiento Estadístico de Patrones</i>	361

<i>Herramientas de análisis de imágenes digitales para identificación y localización de objetos</i>	365
<i>La Realidad Virtual en los Comportamientos Sociales</i>	370
<i>Realidad Aumentada, Realidad Virtual, Interfaces Avanzadas y Juegos Educativos</i>	375
INNOVACIÓN EN EDUCACIÓN EN INFORMÁTICA	
<i>Integridad de datos. Evaluación basada en el método LSP</i>	383
<i>Desarrollo de Aplicaciones Móviles 3D Orientadas a Educación.</i>	388
<i>Aplicando nuevos aspectos en la Programación de Computadoras</i>	392
<i>TIC Sostenibles para la educación y concienciación</i>	397
<i>PROGRAMAR en la Escuela</i>	403
<i>Trabajo colaborativo para el diseño, construcción y recuperación de OA accesibles basados en Realidad Aumentada y Robótica Pedagógica</i>	408
<i>Promoción del pensamiento computacional para el fomento de vocaciones TIC y mitigación de índices de desgranamiento en carreras de Informática</i>	413
INGENIERIA DE SOFTWARE	
<i>Reuso Orientado a Servicios: Selección y Testing de Servicios</i>	418
<i>El modelo de negocio en AOP4ST</i>	423
<i>Reuso Orientado a Dominios: Hacia un Proceso Integral de Desarrollo de Líneas de Productos de Software</i>	428
<i>Software para gestión de requerimientos del Modelo Conceptual de un sistema de información</i>	433
<i>Modelo Integral para la Evaluación de la Calidad de la Accesibilidad al Contenido Web</i>	438
<i>Accesibilidad de la Información en Sitios Web argentinos</i>	443
<i>Un Método para la Evaluación de Modelos Conceptuales de Procesos de Negocio Basado en Lógica Difusa</i>	448
<i>Estudio de los Modelos Conceptuales de Procesos Workflow para Analizar y Evaluar su Migración a la Nube</i>	453
<i>Evaluación de la Usabilidad por medio de Usuarios Finales</i>	458
<i>Mapas Conceptuales para tratar Nominalizaciones en Modelos de Requisitos</i>	463
<i>Técnicas de minería de procesos de negocio distribuidos con Bonita OS</i>	468
<i>Evolución del software: La Accesibilidad Web en Sistemas Gestores de Contenidos de libre distribución</i>	473
<i>Business Intelligence & Knowledge Discovery en el contexto del estándar Project Management Body Of Knowledge del Project Management Institute</i>	478

<i>Definición de Métricas de Calidad para Productos de Software</i>	483
<i>Modelado y Análisis Probabilístico de Sistemas Híbridos</i>	489
<i>Cálculo de Métricas para Medir el Grado de Entendimiento de una Descripción WSDL</i>	493
<i>Procesamiento de Lenguaje Natural para Estudiar Completitud de Requisitos</i>	498
<i>Integración de Metodologías Ágiles y Arquitecturas de Software en el desarrollo de Sistemas de Información</i>	503
<i>Avances en la Construcción de un Sistema de Recuperación de Información para Información Científica en Ciencias de la Computación</i>	508
<i>Marco de trabajo para mejorar el aprovechamiento de factores críticos de éxito en proyectos de ingeniería de software</i>	513
<i>Creación y Simulación de Metodologías de Análisis, Clasificación e Integración de Nuevos Requerimientos a Software Propietario</i>	518
<i>Modelado guiado por arquitecturas de software para un sistema de control de tráfico aéreo en territorio argentino</i>	524
<i>Las TIC al Servicios del Dato Abierto: Situación Actual, Conceptualización e Iniciativas de Apertura de Información Pública</i>	529
<i>Aplicación de principios SOA para la composición de servicios en ambientes ubicuos</i>	534
<i>Aplicaciones Móviles: arquitecturas, visualización, realidad aumentada, herramientas de medición, desarrollo híbrido</i>	539
<i>Extendiendo la meta-arquitectura aportada por el enfoque MDA con conocimiento del dominio</i>	544
<i>Verificación en Alloy de modelos y metamodelos específicos del dominio</i>	549
<i>Formulación de un Modelo de Proceso para Ingeniería del Conocimiento</i>	553
<i>Asistencia a la identificación de especies botánicas del NE del Chubut a través de aplicaciones basadas en ontologías</i>	557
<i>Simulando Proyectos de Desarrollo de Software que incorporan Test Driven Development</i>	562
<i>Análisis de Foros de Discusión para la recuperación de información</i>	567
<i>Dispositivos Móviles: Desarrollo y Análisis de Rendimiento de Aplicaciones Multiplataforma</i>	572
<i>Experiencias en el desarrollo de Sistemas de Software Distribuidos.</i>	576
<i>Guías para aplicación de Normas de Calidad para los procesos de Ingeniería de Software en productos desarrollados con Lenguajes de Programación Open Source: relevamiento y aplicación en PYMES de la zona de influencia de la UNER Concordia</i>	581
<i>Evaluaciones de Accesibilidad y Usabilidad en la WWW: Propuestas para Mejorar la Experiencia del Usuario</i>	586
<i>Metodologías para el desarrollo de software en PYMES</i>	591
<i>Elicitación de requerimientos con Grounded Theory</i>	596
<i>Aplicación de patrones en el modelado de procesos de negocio</i>	601
<i>Aplicación de tecnologías semánticas a la Forensia Digital. Etapa 1: Estudio y Diseño de una Ontología Semántica</i>	606

<i>Gestión del Conocimiento Abierto</i>	611
<i>Aseguramiento de la calidad en productos, procesos de software y procesos de gestión para la mejora de las sociedades del conocimiento</i>	616
<i>MoVeR-T: Entorno de videojuego en Realidad Virtual Telecontrolado para Rehabilitación Motriz</i>	621
<i>Mejoras en la ejecución de BPM incluyendo conceptos de Green IT</i>	626
<i>Estudio de Adopción de Técnicas de Desarrollo de Software Guiado por las Pruebas</i>	631
<i>Mejora y gestión de procesos de negocio inter-organizacionales aplicando técnicas de minería de procesos</i>	637
<i>Especificación de indicadores en el proceso de software con tecnologías libres</i>	641
<i>Análisis sobre adopción de metodologías ágiles en los equipos de desarrollo en pymes del NEA</i>	646
<i>Métodos y herramientas de estimación, gestión cuantitativa de proyectos, trazabilidad de requerimientos y entrega continua, orientados a la mejora de la calidad del software</i>	651
<i>Framework de Evaluación de Productos Software</i>	657

INNOVACIÓN EN SISTEMAS DE SOFTWARE

<i>Implementación de Sistemas Inteligentes para la Asistencia a Alumnos y Docentes de la Carrera de Ingeniería en Sistemas de Información</i>	662
<i>Realidad Aumentada Aplicada a la Asistencia Médica en el Campo de la Emergentología</i>	667
<i>Informatización de datos Botánicos de la Universidad Nacional de Luján: un camino al Sistema de Bioinformática de la Institución</i>	672
<i>Estudio, análisis y diseño de estrategias de diseño de software colaborativo</i>	677
<i>Implementación de un sistema de video conferencia HD autónomo utilizando Raspberry Pi</i>	682
<i>Gestión, Inventario y Monitoreo Hardware con Alertas Automáticas (G.I.M.H.A.A.)</i>	687
<i>Una Arquitectura Cliente-Servidor para Modelado Conceptual Asistido por Razonamiento Automático</i>	692
<i>Aplicación móvil con finalidad de ahorro sobre compras en supermercados.</i>	697
<i>Linked Open Data para la Integración de Información Científica</i>	701
<i>La gestión de la información en abierto, vehículo importante para maximizar la visibilidad web</i>	706
<i>Diseño de Sistema IoT de monitoreo y alarma para personas mayores</i>	712
<i>La Usabilidad a través de Modelos Abstractos empleando el Desarrollo de Software dirigido por Modelos</i>	717
<i>Mejora de la precisión posicional utilizando receptores GPS de bajo costo</i>	722
<i>Comunicación alternativa y aumentativa para potenciar la autonomía personal y la calidad de vida de las personas con discapacidades severas</i>	726

<i>Bioestadística Aplicada: Software DL50</i>	732
---	-----

<i>Living Labs en la región Noroeste de la provincia de Buenos Aires</i>	738
--	-----

PROCESAMIENTO DISTRIBUIDO Y PARALELO

<i>Performance de arquitecturas multiprocesador: técnicas de modelado y simulación y plataformas reconfigurables</i>	743
--	-----

<i>Performance de cloud computing para HPC: despliegue y simulación</i>	748
---	-----

<i>Evaluación de arquitecturas distribuidas de commodity basadas en software libre</i>	752
--	-----

<i>Paralelización de Aplicaciones Económicas que requieren Estimación de los Modelos de Elección Discreta</i>	757
---	-----

<i>Cluster de microprocesadores RISC para problemas de agricultura de precisión</i>	762
---	-----

<i>Identificación de Algoritmos de Cómputo Intensivo para Big Data y su Implementación en Clouds</i>	767
--	-----

<i>Fundamentos de cómputo paralelo y distribuido para HPC. Construcción y evaluación de aplicaciones.</i>	772
---	-----

<i>Evolución y Tendencias en Sistemas Paralelos para HPC</i>	778
--	-----

<i>Arquitecturas Multiprocesador en HPC: Software, Métricas y Aplicaciones.</i>	784
---	-----

<i>Computación de Altas Prestaciones: Problemáticas y Aplicaciones</i>	789
--	-----

<i>Computación de Alto Desempeño y Datos Masivos: Arquitecturas, Modelos y Paradigmas.</i>	794
--	-----

PROCESAMIENTO DE SEÑALES Y SISTEMAS DE TIEMPO REAL

<i>Estudio y Desarrollo de Interfaces Avanzadas orientadas a Sistemas de Robótica</i>	799
---	-----

<i>Análisis de Señales Acústicas para dispositivos de personas con discapacidad, aplicando Teoría de Procesamiento de Radar</i>	805
---	-----

<i>Control programático de Drone DJI Phantom 3</i>	811
--	-----

<i>Avances en robótica situada aplica a la navegaciones autónoma de cuadricópteros</i>	816
--	-----

<i>Sistemas de Tiempo Real Mixtos: Planificación en Sistemas Operativos de Tiempo Real Bajo Plataformas de Desarrollo Concretas</i>	821
---	-----

<i>Sistemas de Tiempo Real Distribuidos Robots y Microcontroladores</i>	826
---	-----

<i>Sintonía de Controladores Inteligentes Mediante Estrategia Híbrida Fuzzy-PSO</i>	830
---	-----

<i>Robótica y Fenotipado de Alta Capacidad con Relevamiento de Datos en Campo. Aplicaciones en Agricultura de Precisión</i>	837
---	-----

<i>Planificación de tareas en Sistemas de Tiempo-Real</i>	842
---	-----

SEGURIDAD INFORMÁTICA	
<i>Un Modelo para la Evaluación de la Seguridad en Sistemas Informáticos</i>	845
<i>Análisis comparativo de Algoritmos Criptográficos Livianos para dispositivos RFID de bajo costo</i>	850
<i>Anonimato en Sistemas de e-Voting: Últimos Avances</i>	854
<i>Esteganografía: Sustitución LSB 1 bit utilizando Matlab</i>	859
<i>Secuencias Seudoaleatorias para Criptología</i>	865
<i>Optimización de las fórmulas para la detección de Infraestructuras de Clave Pública anómalas</i>	868
<i>Avances en Educación de Dinámica de Tecleo y el Contexto Emocional de un Individuo aplicando Interfaz Cerebro Computadora</i>	872
<i>Estructuras algebraicas aplicables en criptografía</i>	877
<i>Ciberdefensa en redes industriales</i>	882
<i>Modelo de Análisis Relativo a la Protección de Datos Personales para Proyectos de Cómputo en la Nube</i>	886
<i>Análisis digital forense utilizando herramientas de software libre</i>	891
TECNOLOGÍA INFORMÁTICA APLICADA EN EDUCACIÓN	
<i>Una evaluación de los simuladores de autómatas existentes a los fines de su implementación práctica en la materia Sintaxis y Semántica de los lenguajes en la carrera de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Córdoba</i>	895
<i>Herramientas Colaborativas Multiplataforma en la Enseñanza de la Ingeniería de Software</i>	899
<i>Plataforma de Gamificación para la Enseñanza de Programación en el Ámbito Universitario</i>	905
<i>Pruebas y materiales instruccionales en ambientes virtuales de aprendizaje adaptativos</i>	910
<i>Entornos virtuales en la enseñanza de las Ciencias Exactas y Naturales</i>	915
<i>Proyecto: Inteligencia Artificial y desarrollo de Simuladores hacia el diseño de Cursos Abiertos On Line</i>	919
<i>Estudio de la influencia del uso de simulación en la enseñanza de redes de computadoras en el nivel universitario (resultados parciales)</i>	924
<i>Técnicas provenientes de las ciencias de la computación aplicadas a la simulación macroeconómica</i>	931
<i>Evaluación de aprendizajes: instrumentos e interactividad en entornos virtuales</i>	935
<i>Hacia un modelo predictivo de rendimiento académico utilizando minería de datos en la UTN – FRRe</i>	940
<i>TIC y Objetos de aprendizaje en el ámbito educativo</i>	945
<i>Aplicaciones para la enseñanza aprendizaje de matemática en el nivel secundario</i>	951

<i>Software educativo para la resolución numérica y gráfica de ecuaciones diferenciales ordinarias</i>	956
<i>Realidad Aumentada: innovación educativa en las aulas jujeñas</i>	960
<i>Propuesta de un Marco Conceptual para el Diseño e Implementación de Repositorios Institucionales de Contexto Educativo Ubicuos (RICEU)</i>	965
<i>M-library: Servicios de una biblioteca universitaria disponibles a través de dispositivos móviles</i>	970
<i>Programa de I+D+I en Ingeniería de Espacios Virtuales de Trabajo</i>	975
<i>Diseño y Desarrollo de Aplicaciones de Aprendizaje Ubicuo</i>	978
<i>Escenarios educativos mediados por Tecnologías de la Información y la Comunicación</i>	983
<i>Computación Aplicada: Búsqueda y Desarrollo de Nuevas Estrategias de Enseñanza y Recursos TIC</i>	989
<i>Metabuscarador basado en agentes para grupos de estudiantes colaborativos</i>	994
<i>Aplicaciones basadas en ontologías para asistir a la integración del aprendizaje móvil en escuelas de Catamarca</i>	998
<i>Ontologías para entornos personalizados de aprendizaje ajustados al contexto educativo argentino</i>	1003
<i>RedAUTI Red de Aplicaciones y Usabilidad de la TVDi</i>	1008
<i>Selección y manipulación de Recursos Educativos Abiertos en el entorno educativo argentino mediante aplicaciones basadas en ontologías</i>	1015
<i>Paradigmas de Interacción Persona Ordenador en el ámbito de la Educación y la Educación Especial. Avances del proyecto y resultados</i>	1020
<i>La tecnología como vehículo de articulación Nivel Medio / Universidad en la provincia de La Rioja</i>	1026
<i>Caracterización del uso de Aulas Virtuales en el Ciclo de Formación Específica de Carreras de Ingeniería</i>	1032
<i>Implicancias conceptuales interdisciplinarias en la búsqueda de creatividad y criticidad para ambientes colaborativos y distribuidos soportados por computador y mediados por la gestión del conocimiento</i>	1037
<i>Elementos interactivos en los mundos virtuales 3 D</i>	1041
<i>Herramientas informáticas orientadas a la enseñanza y el aprendizaje</i>	1045
<i>Prácticas de m-learning desde un enfoque ecológico</i>	1050
TESIS DOCTORALES	
<i>Desarrollo Dirigido por Modelos Basado en Componentes de Interfaz de Usuario</i>	1055
<i>Representación y razonamiento sobre las decisiones de diseño de arquitectura de software.</i>	1065
<i>Marcos Argumentativos Etiquetados</i>	1075
<i>Consolidación de Ontologías Datalog+/-</i>	1085

Utilización de técnicas subsimbólicas de la inteligencia artificial para la generación de energías limpias

D. Martín Morales^{1,2}, Marcelo A. Cappelletti^{1,3},
Waldo Hasperué^{1,4}, Leandro J. Charlier¹

¹Instituto de Ingeniería y Agronomía,
Universidad Nacional Arturo Jauretche (UNAJ)

²Universidad Tecnológica Nacional, Facultad Regional La Plata (UTN, FRLP)

³Grupo de Estudio de Materiales y Dispositivos Electrónicos (GEMyDE) -
Fac. Ingeniería, Universidad Nacional de La Plata (UNLP)

⁴Instituto de Investigación en Informática LIDI (III-LIDI) – Fac. Informática, UNLP

martin.morales@unaj.edu.ar, mcappelletti@unaj.edu.ar,
whasperue@unaj.edu.ar, leandrocharlier@gmail.com

Resumen

La contaminación ambiental acumulada durante las últimas décadas sumado a la constante disminución de los recursos naturales de fuentes de energía no renovables, hacen que la conversión de energía sea uno de los problemas más importantes a resolver a nivel mundial. Sobre este punto, la línea de investigación que se presenta se basa en la utilización de técnicas de inteligencia artificial subsimbólica para el estudio de sistemas físicos relacionados con aplicaciones ambientales para la generación de energías limpias. En particular, se propone el estudio de celdas solares multijunturas para aplicaciones terrestres y espaciales, con el propósito de obtener diseños óptimos de estos dispositivos con una relación costo-eficiencia por encima de las obtenidas actualmente. Estos dispositivos constituyen una de las tecnologías más prometedoras para generar energía de forma no contaminante

y eficiente para aplicaciones domésticas, industriales y en vehículos espaciales.

La utilización de técnicas subsimbólicas de la inteligencia artificial como instrumentos de cálculo y análisis, posibilita reproducir resultados experimentales, realizar estudios sobre sistemas, procesos, propiedades o parámetros físicos difíciles de medir de manera experimental, permitiendo reducir los tiempos y los costos de los ensayos.

Palabras clave: *Inteligencia artificial subsimbólica. Celdas solares multijunturas. Modelización. Simulación. Diseño y optimización.*

Contexto

Se presenta una línea de Investigación que es parte del Proyecto de Investigación Orientado PIO CONICET-UNAJ 2015-2016, denominado “Utilización de

métodos numéricos y de técnicas subsimbólicas de la inteligencia artificial para la generación de energías limpias y la evaluación de la salud ósea”, de la Universidad Nacional Arturo Jauretche (UNAJ). Se trata de un proyecto que se encuentra en su etapa inicial y cuenta con financiamiento de CONICET.

El proyecto tiene como objetivo el estudio de sistemas físicos relacionados con aplicaciones ambientales para la generación de energías limpias y con aplicaciones biomédicas para la evaluación de la salud ósea. Estos sistemas representan un gran impacto en el desarrollo tecnológico y en la mejora de la calidad de vida de la humanidad.

En el tema existe un convenio de colaboración en actividades de Investigación y Postgrado con el Grupo de Estudio de Materiales y Dispositivos Electrónicos (GEMyDE) de la Facultad de Ingeniería de la Universidad Nacional de La Plata (UNLP).

Introducción

Inteligencia Artificial Subsimbólica

El comportamiento de un determinado sistema físico puede ser descrito a partir de modelos matemáticos formulados a través de ecuaciones diferenciales con condiciones de contorno e iniciales determinadas, cuyas incógnitas representan las magnitudes de interés. Tales modelos representan generalmente una simplificación de la realidad, dado que son obtenidos a partir de aproximaciones, sin embargo, si las variables a despreciar tienen poco peso en el sistema, serán una buena representación del sistema real.

Obtener la solución exacta de manera analítica puede presentar una gran dificultad, debido a la complejidad

natural de las ecuaciones diferenciales o a los inconvenientes que pueden ocurrir con las condiciones de contorno o iniciales.

En los últimos años, el crecimiento de las capacidades de cálculo y el mejoramiento de los algoritmos implementados han convertido a la simulación numérica en una etapa indispensable a la hora de estudiar cualquier tipo de sistema.

En este sentido, las técnicas subsimbólicas de la inteligencia artificial constituyen herramientas poderosas para resolver tales ecuaciones y obtener soluciones muy aproximadas a las soluciones exactas. De esta manera, es posible reproducir resultados experimentales; realizar estudios sobre sistemas, procesos, propiedades o parámetros físicos difíciles de medir de manera experimental reduciendo los tiempos y los costos de los ensayos; analizar, predecir y controlar la respuesta de un sistema determinado bajo diferentes condiciones de operación, con el fin de diseñar el sistema de manera eficiente para obtener la respuesta deseada de una aplicación específica; etc.

El campo denominado Inteligencia Artificial Subsimbólica (AI: Artificial Intelligence) [1], incluye técnicas tales como: Redes Neuronales Artificiales (ANN: Artificial Neural Network), Lógica Difusa (FL: Fuzzy Logic), Sistema de Inferencia Difuso basado en Redes Adaptativas (ANFIS: Adaptive Network based Fuzzy Inference System), Algoritmos Genéticos (GA: Genetic Algorithms), Optimización por Enjambre de Partículas (PSO: Particle Swarm Optimization) y Minería de Datos (DM: Data Mining). Una de las características de estas técnicas es que son del tipo caja negra, es decir que no permiten obtener una función explícita de las variables físicas involucradas. Sin embargo, ellas tienen el potencial para

realizar mejores, más rápidas y más prácticas predicciones que cualquier otro método tradicional. Son capaces de encontrar correlaciones entre datos diversos y han mostrado ser sumamente útiles en áreas tales como ingeniería, análisis de suelos, análisis climático, ecología, análisis genético, robótica, economía, medicina y comunicación, entre otras.

En particular, recientemente, técnicas subsimbólicas de inteligencia artificial se están utilizando cada vez más en aplicaciones relacionadas con la generación de energía eléctrica a través de fuentes renovables de energía, por ejemplo para la determinación de parámetros físicos de celdas solares [2-4].

Celdas solares multijunturas

El gran potencial de energía solar disponible se presenta como un recurso muy importante tendiente a sustituir las fuentes de energía no renovables (carbón, petróleo, gas natural) y reducir las emisiones de dióxido de carbono (CO₂) a la atmósfera, con el propósito de proteger nuestro medio ambiente y mejorar la calidad de vida de la humanidad. En este sentido, las celdas solares para la generación de energía eléctrica, son desarrollos cada vez más prometedores. Estos dispositivos convierten la luz solar directamente en electricidad sin pasar por un ciclo térmico, es decir que pueden proveer potencia casi permanentemente a un bajo costo operativo. Es una forma "limpia" de producir electricidad dado que la interacción de las celdas solares con el medio ambiente, en cuanto a los daños que puedan ocasionar, es considerablemente menor que la producida por los generadores de energía convencionales utilizados hoy en día.

Sin embargo, desafortunadamente aún hoy la energía obtenida de las celdas

solares resulta costosa. Todavía existe una enorme brecha entre el potencial de energía solar disponible y su real aprovechamiento, debido principalmente a la modesta densidad de energía de la radiación, las bajas eficiencias de conversión de las celdas solares y al costo de los materiales requeridos. Mejorar la relación costo-efectividad es una tarea que despierta gran interés y es todo un desafío científico.

En la actualidad, casi el 90% de las celdas solares disponibles en el mercado mundial están basadas en la tecnología de silicio mono- y poli-cristalino, la cual consiste en dispositivos de simples junturas p-n. Si bien es una tecnología probada, confiable y de larga vida útil, ya se han logrado eficiencias máximas cercanas al 25%, valor muy próximo a su límite teórico calculado por Shockley-Queisser en base a argumentos de balance detallado de energía, de 31% bajo 1 sol de iluminación y 40.8% bajo la máxima concentración de soles [5]. La razón fundamental de este límite es que sólo los fotones incidentes con una energía superior a la de la banda prohibida del semiconductor (bandgap) pueden liberar un electrón que contribuirá a la corriente eléctrica. Los fotones con energía menor se pierden (el semiconductor es transparente para ellos).

Durante los últimos años, nuevos conceptos han permitido superar las eficiencias mencionadas previamente, a partir del desarrollo de celdas solares basadas en nuevos materiales fotovoltaicos y nuevos procesos de fabricación. Una de ellas es mediante la utilización de dos o más celdas solares, apiladas una encima de la otra, subdividiendo el espectro solar en diferentes rangos de energía. Con esta idea es que surgen las celdas solares tipo tandem o multijunturas (MJSC: Multi-Junction Solar Cells), las cuales se

forman a partir de dos o más sub-celdas de materiales con diferente bandgap, separadas entre sí por juntas tipo túnel, para facilitar el flujo de electrones entre ellas. Las sub-celdas son colocadas en orden descendente de acuerdo a su bandgap. La sub-celda ubicada en la parte frontal del dispositivo captura los fotones de más alta energía y deja pasar el resto de los fotones hacia abajo para ser absorbidos por las sub-celdas con menores bandgap. De esta manera, cada sub-celda absorbe una porción diferente del espectro solar, mejorando el rendimiento en comparación con las celdas individuales separadas. Estos dispositivos tienen entre sus ventajas una alta sensibilidad en un amplio rango de longitudes de onda y un excelente rendimiento. En el límite teórico, cuando la MJSC esté compuesta por un número infinito de sub-celdas, los cálculos establecen una eficiencia máxima del 68.2% y del 86.8% bajo 1 sol y bajo la máxima concentración de soles, respectivamente [6], muy superior a las celdas de simple junta, lo que implica que más juntas aumentan su eficiencia pero al mismo tiempo se incrementan rápidamente tanto el costo como su complejidad de fabricación.

En particular, las MJSC basadas en semiconductores compuestos III-V representan hoy en día el estado del arte de la tecnología solar fotovoltaica (por ejemplo: GaInP/GaInAs/Ge o GaInP/GaAs/Ge). Recientemente han sido obtenidas eficiencias de conversión record por encima del 41 % para 454 soles de iluminación [7].

En consecuencia, con esta línea de investigación del proyecto, pretendemos contribuir al estudio de las MJSC con diferentes características constructivas y diferentes condiciones de operación, tanto para aplicaciones terrestres como espaciales, con el fin de obtener una

relación costo-eficiencia mejorada de las mismas. El estudio de estos dispositivos requiere de un profundo conocimiento de los aspectos teóricos concernientes a sus propiedades de interés y a los parámetros que gobiernan los cambios de dichas propiedades.

Líneas de Investigación, Desarrollo e Innovación

- Estudio de diferentes métodos de inteligencia artificial subsimbólica aplicados al análisis de dispositivos fotovoltaicos, considerando ventajas y desventajas de cada uno, viabilidad y factibilidad de implementación, procedimiento, etc..
- Utilización de técnicas subsimbólicas de la inteligencia artificial para el análisis de MJSC basadas en heterojuntas de semiconductores compuestos III-V, variando los principales parámetros a ser considerados durante la fase de diseño de los dispositivos: densidades de portadores, espesores de las regiones que lo conforman, materiales, tipo y número de juntas, y condiciones de operación, tanto para aplicaciones terrestres como espaciales, con la finalidad de diseñar y optimizar dispositivos con mayores eficiencias y tolerancia a la radiación.

Resultados y Objetivos

- Determinar, a través de técnicas subsimbólicas de la inteligencia artificial, parámetros físicos no conocidos de las celdas solares multijuntas y de las sub-celdas que las componen, basadas en diferentes materiales, estructuras, y condiciones

de operación, tanto para aplicaciones terrestres como espaciales.

- Obtener modelos analíticos de gran utilidad para el diseño de dispositivos con características específicas para optimizar la relación costo-beneficio;
- Predecir el comportamiento de la celda solar para diferentes temperaturas e irradiancias;
- Fomentar la generación de recursos humanos en esta línea de investigación, donde se conjuga la física de los dispositivos, y la informática a través del desarrollo de herramientas computacionales.

Formación de Recursos Humanos

El equipo de trabajo de la línea de investigación que se presenta está conformado por tres docentes investigadores (Morales, Cappelletti y Hasperué) y por un estudiante de quinto año de la carrera Ingeniería en Informática de la UNAJ (Charlier), autores de este artículo. Actualmente se encuentran en curso una Tesis de Posgrado (Morales) y una Tesina de Grado (Charlier) relacionadas directamente con la línea de investigación presentada.

Dado que se trata de una línea de Investigación con características multidisciplinarias, se abre una puerta para la inclusión de docentes investigadores y estudiantes que deseen desarrollarse en estas aplicaciones relacionadas con la informática y el medio ambiente.

Referencias

[1] Nilsson N., "Principles of Artificial Intelligence", Springer-Verlag, ISBN 3-540-11340-1, New York (1980).

[2] Dkhichi F. et al, "Determination of Solar Cell Parameters using Neural Network Trained by Steepest Descent Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, 4 pp. 121-125 (2014).

[3] Ospino Castro A. et al, "Modelado y simulación de un panel fotovoltaico empleando técnicas de inteligencia artificial", Ingeniería Energética Vol. XXXV, pp.225-233 (2014).

[4] Salem F. et al, "Parameters estimation of photovoltaic modules: comparison of ANN and ANFIS", Int. J. Industrial Electronics and Drives, 1 pp. 121-129 (2014).

[5] Shockley W., Queisser H.J., Journal of Applied Physics, 32, pp. 510-519 (1961).

[6] Yastrebova N.V., Centre for Research in Photonics, University of Ottawa (2007).

[7] Guter W. et al., "Current-matched triple-junction solar cell reaching 41.1% conversion efficiency under concentrated sunlight", Applied Physics Letters 94, pp. 223504 (2009).

Operadores de Cambio Múltiples en Bases de Creencias Bajo Lógica Horn

Néstor Jorge Valdez[†]

Marcelo A. Falappa[‡]

[†] Departamento de Ciencias de la Computación, Fac. de Ciencias Exactas y Naturales
Universidad Nacional de Catamarca (UNCa)
Av. Belgrano 300 - San Fernando del Valle de Catamarca
Tel.: (03834)420900 / Cel: (03834) 154066788
e-mail: njvaldez@c.exactas.unca.edu.ar

[‡] Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina
Tel: (0291)4595135 / Fax: (0291)4595136
e-mail: mfalappa@cs.uns.edu.ar

Resumen

Esta línea de investigación tiene como objetivo estudiar en el marco de la teoría de cambio de creencias los operadores de cambio múltiple, es decir, operadores en que la entrada epistémica es un conjunto de sentencias en lugar de una única sentencia. Para ello, se analizan los resultados generales presentados sobre operadores de *revisión múltiples priorizadas y no-priorizadas* en bases de creencias y otros referidos a modelos de contracción simultánea para conjunto de sentencias denominada *package contraction*. Estudiaremos y definiremos a estos operadores de cambios múltiple bajo una lógica más restringida que la lógica proposicional clásica como es la *lógica Horn*. Así también, investigar diferentes operadores de cambio y las interrelaciones entre ellos. Esto permitirá desarrollar nuevos algoritmos de cambio múltiple y sus respectivas caracterizaciones axiomáticas.

Palabras Claves Revisión de Creencias,

Cambio Múltiples Horn, Bases de Creencias.

1. Contexto

Esta línea de investigación se realizará dentro del ámbito del Laboratorio de Investigación del Departamento de Ciencias de la Computación de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Catamarca. Esta investigación forma parte de las contribuciones de la tesis para la obtención del título de Doctor en Ciencias de la Computación de la Universidad Nacional del Sur por parte de uno de los autores del presente trabajo: Néstor Jorge Valdez. Así también, esta temática de estudio está asociado con el proyecto de investigación Bianual: “*Cambio de Creencias Múltiples en Sistemas Argumentativos Aplicado para Programación Lógica Rebatible*”, financiado por el programa de desarrollo científico y tecnológico de la Secretaría de Ciencia

y Tecnología: Consejo de Investigación, de la Universidad Nacional de Catamarca. Periodo: 01/01/2016 al 31/12/2017.

2. Introducción

Uno de los desafíos que enfrenta la lógica de la teoría de cambio consiste en el problema de modelar la dinámica de conocimiento. Un sistema de revisión de creencias constituye un marco lógico para modelar la dinámica de conocimiento, esto es, cómo modificamos nuestro estado epistémico cuando recibimos información nueva. Generalmente, la nueva información es asumida mediante una sentencia simple, como en la teoría AGM clásica [1, 2], pero la entrada epistémica también podría presentarse a través de un conjunto de sentencias en lo que se conoce como cambio múltiple.

En este contexto, existen diversos tipos de operadores de cambio, como los operadores de *cambio priorizado*, los operadores *no priorizados*, los operadores *merging* o *cambio simétrico*, los operadores de *cambios selectivos*, y los operadores de *semi-revisión*, entre otros [6].

Debido a que las operaciones de cambio se llevan a cabo sobre el estado epistémico del agente, existen diferentes posibles construcciones para los distintos operadores de cambio. Dos de los modelos más comúnmente usados son sobre *conjuntos de creencias* y sobre *bases de creencias*. En general, nuestras propuestas están centradas en el uso de bases de creencias puesto que si contamos con un lenguaje (al menos proposicional) lo suficientemente declarativo, no es posible tratar con conjuntos clausurados. Sabemos que el modelo AGM asume una lógica subyacente que es al menos tan expresiva como la lógica proposicional. Debido a este supuesto, esta teoría no puede ser aplicada a los sistemas con lógicas subyacentes decla-

rativamente más ricas que la lógica proposicional como por ejemplo aquellos sistemas de Inteligencia Artificial que utilizan lenguajes de primer orden o algún subconjunto de ellos. Entre estos lenguajes, por cualidades computacionales ampliamente conocidas, es que utilizaremos *lógicas Horn*.

Con respecto esto último, los primeros resultados en cambio de creencias Horn fueron obtenidos por James Delgrande [3]. Diversas investigaciones demuestran utilizar la lógica Horn como lenguaje subyacente permite la posibilidad de más de un tipo de contracción. En cuanto a una representación de resultados para el tipo de contracción múltiple denominada *package contraction* bajo lógicas Horn, existen diversas maneras en que la contracción Horn (*simple* y *múltiple*) se diferencia de la contracción AGM. Una de ellas es la ausencia del postulado de *recovery*. Esto se debe a que, al menos para cláusulas Horn, el postulado *recovery* resulta inapropiado. Otra manera es que en la contracción Horn el resultado de la misma puede expresarse en función de un conjunto de restos. En cambio en la contracción AGM, un único conjunto de restos lleva a las funciones a realizar cambios, pero con propiedades no deseables por tratar con conjuntos clausurados.

3. Línea de Investigación y Desarrollo

Esta línea de investigación toma como punto de partida publicaciones en donde se presentan los aportes realizados en la temática de los operadores de contracción múltiple de la teoría de cambio de creencias bajo cláusulas Horn que han sido investigado en trabajos previos por los autores de este artículo [9, 10, 11, 12]. Uno de los principales intereses es estudiar las operaciones de cambio, en particular aquellos cambios de

creencias que son inducidos a través de un conjunto (múltiple) de sentencias. Para ello, consideramos principalmente dos tipos diferentes de cambio múltiple. El primero abarcará operadores de cambio múltiple priorizado en donde todas las nuevas creencias se supone que deben ser aceptadas, y el segundo se referirá a operadores de mezcla (merging) que permite que creencias antiguas y nuevas jueguen roles simétricos dentro de un proceso de cambio. Para cada tipo de cambio, analizamos dos construcciones conocidas: uno basada en kernels y otra basada en conjuntos de restos.

El término 'revisión múltiple' se utiliza para referirse a las operaciones de revisión que permite la revisión simultánea por más de una sentencia. Distinguimos este tipo de operaciones de las operaciones *repetidas* o *iteradas*, en los cuales se aplica una secuencia de dos o más cambios-. Al considerar la revisión múltiple priorizada, podríamos pensar en hacer uso de la revisión AGM clásica (es decir, la revisión por una sola sentencia) a través de la aplicación de la negación sentencial.

Sea \div un operador de contracción AGM para K . Una forma de definir una revisión múltiple es mediante el uso de la *Identidad de Levi Generalizado*:

$$K * A = (K \div n(A)) \cup A$$

donde A es un conjunto, \div es un operador de contracción AGM clásico y $n(A)$ es la negación sentencial de A , la cual se define como sigue [8, 7]:

1. $n(\emptyset) = \perp$.
2. Si A es un conjunto unitario ($A = \{\alpha\}$), entonces $n(A) = \neg\alpha$.
3. Si $A = \{\alpha_1, \dots, \alpha_m\}$ para algún $m > 1$, entonces $n(A) = \neg\alpha_1 \vee \dots \vee \neg\alpha_m$.

La caracterización axiomática podría obtenerse a partir de una adaptación de las caracterizaciones correspondientes de los modelos tradicionales. Como uno de los objetivos que proponemos para esta investigación es estudiar las interrelaciones entre los operadores de revisión múltiple y las distintas clases de operadores de contracción múltiples de las denominadas *package contraction*, es que resulta importante considerar las versiones generalizadas de la Identidad de Levi.

Sin embargo, no partiremos esencialmente de la negación sentencial al revisar un conjunto arbitrario K con respecto a otro conjunto arbitrario A , si no que consideraremos dos tipos de operadores directamente sin definirlos a partir de sendas operaciones de contracción. En particular, definiremos dos tipos de revisiones priorizadas, adaptadas de [6]:

- Revisión Kernel Múltiple.
- Revisión Partial Meet Múltiple.

La primera construcción de revisión múltiple por un conjunto de sentencias está basado en el concepto de un *A-inconsistent-kernels*. Aquí, hay que definir una función de incisión que realiza un corte en cada *inconsistent-kernel*. Más precisamente, dicha función debe ser una *función de incisión consolidada*, en el sentido de que todas las sentencias de A están protegidas y no pueden ser removidas por esta función. Esto es, una función de incisión consolidada selecciona entre las sentencias de $K \setminus A$ que hace $K \cup A$ inconsistente.

La segunda construcción de revisión múltiple por un conjunto de sentencias está basado en el concepto de un *A-consistent-remainders*. En este caso, se define una función de selección que selecciona los 'mejores' *consistent-remainders*. La función debe ser una *función de selección consolidada*, en el

sentido de que todas las sentencias de $K \cap A$ están protegidas, y ellas están incluidas en la intersección de algún conjunto de restos. Luego, una función de selección consolidada selecciona un subconjunto del conjunto de $K \perp_{\top} A$ cuyos elementos (todos) contienen el conjunto $K \cap A$.

A diferencia de la preferencia absoluta para con la nueva información implícita en el modelo AGM, podemos cambiar nuestras creencias de una manera *no priorizada*. Entre los operadores más conocidos sobre cambio no priorizado en conjuntos de creencias están: *credibility limited revision*, *selective revision* y *screened revision*; en lo que respecta a bases de creencias los operadores sobre cambio no priorizado podemos citar los operadores de cambio basados en *explicaciones* [4, 5]. Nuestra investigación también estará centrada en un tipo especial de cambio no priorizado en bases de creencias llamadas *merging* [6].

Merging representa un problema que ocurre en diversas situaciones, por ejemplo, en el campo de las ciencias informáticas en las que es necesario sintetizar adecuadamente varias fuentes de información. Esta operación para fuentes múltiples de información se puede aplicar en bases de datos distribuidas, sistemas multi-agente, sistemas expertos, etc. En estas situaciones, es importante integrar múltiples bases de datos en una base de datos única y preferiblemente consistente. Una de las aplicaciones más importantes de merging está en la integración de sistemas expertos, donde cada uno tiene una base de creencias que representa su estado de conocimiento. Así también, los estados de creencias se pueden representar con un lenguaje proposicional; en otros casos, puede ser necesario un subconjunto de una lógica de primer orden como por ejemplo, programas lógicos o Horn. En cada caso, consideraremos también dos clases de operadores: basados en conjuntos kernels y en conjuntos

de restos.

4. Resultados y Objetivos Esperados

A través de esta línea de investigación se propone estudiar y definir nuevas construcciones de operadores de revisión múltiples priorizados y no priorizados. Para cada nueva construcción se pretende establecer un conjunto de postulados que caractericen axiomáticamente a la operación de cambio y determinar relaciones entre las construcciones y los postulados. Se busca, además, permitir una eficiente implementación de estos nuevos algoritmos. Analizar a estos operadores de cambios múltiples bajo una lógica más restringida que la lógica proposicional clásica como es la *lógica Horn*. Esto, permitirá desarrollar nuevos procedimientos algorítmicos de cambios múltiples buscando alcanzar caracterizaciones axiomáticas e implementaciones eficientes.

5. Formación de Recursos Humanos

El Docente-Investigador Néstor Jorge Valdez es dependiente del Programa de Desarrollo Científico y Tecnológico (adherido al Programa Nacional de Incentivos a los Docentes-Investigadores) de la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Catamarca, es Magister en Ciencias de la Computación egresado de la Universidad Nacional del Sur, y aspira a alcanzar el título de Doctor en Ciencias de la Computación en esta última casa de estudios.

Referencias

- [1] Alchourrón, Gärdenfors, and Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50:510–530, 1985.
- [2] Alchourrón and Makinson. On the logic of theory change: Contraction functions and their associated revision functions. *The Journal of Symbolic Logic*, 48:14–37, 1982.
- [3] Delgrande. Horn clause belief change: Contraction functions. In *Gerhard Brewka and Jérôme Lang, editors, Proceedings of the Eleventh International Conference on the Principles of Knowledge Representation and Reasoning, Sydney, Australia, 2008*. AAAI Press, pages 156–165, 2008.
- [4] Falappa, Kern-Isberner, and Simari. Belief revision, explanations and defeasible reasoning. *Artificial Intelligence Journal*, 141:1–28, 2002.
- [5] Falappa, Kern-Isberner, and Simari. Argumentation in artificial intelligence. In *I. Rahwan, G. R. Simari (Eds.), Belief revision and Argumentation Theory*, pages 341–360, 2009.
- [6] Falappa, Kern-Isberner, and Simari. Prioritized and non-prioritized multiple change on belief bases. *Journal of Philosophical Logic, ISSN 0022-3611, New York: Springer*, 41:77–113, 2012.
- [7] Peter Gärdenfors. *Belief Revision*. Cambridge University Press, 1992.
- [8] Sven Ove Hansson. A dyadic representation of belief. In Gardenfors [7], pages 89–121.
- [9] Valdez and Falappa. Dinámica de conocimiento: Contracción múltiple en lenguajes horn. *XIX Congreso Argentino de Ciencias de la Computación, XIV Workshop Agentes y Sistemas Inteligentes (WASI), CACiC'2013*, pages–, 2013.
- [10] Valdez and Falappa. Dinámica de conocimiento: Contracciones horn a partir de ordenamientos epistémicos. *42JAIIO Jornadas Argentinas de Informáticas, ASAI 2013, 42 JAIIO'2013*, pages 206–209, 2013.
- [11] Valdez and Falappa. Implementación para bases de creencias horn de operadores de contracción múltiple. *XX Congreso Argentino de Ciencias de la Computación, XV Workshop Agentes y Sistemas Inteligentes (WASI), CACiC'2014*, 2014.
- [12] Valdez, Lara, Pedraza, and Teseira. Dinámica de conocimiento: Cambio de creencias múltiples. temática de investigación de inteligencia artificial en las ciencias de la computación. *JUCEN'15, Facultad de Ciencias Exactas y Naturales, UNCa. Catamarca 9 y 10 de diciembre de 2015 - Argentina*, pages 30–36, 2015.

Cambio de Creencias Múltiples en Sistemas Argumentativos Aplicado para Programación Lógica Rebatible

Luis Gustavo Pedraza[†] Oscar Andrada[†]
Néstor Jorge Valdez[†] Mario Guido Sergio Teseira[†]

[†] Departamento de Ciencias de la Computación, Fac. de Ciencias Exactas y Naturales
Universidad Nacional de Catamarca (UNCa)
Av. Belgrano 300 - San Fernando del Valle de Catamarca
Tel.: (03834)420900
e-mails: lgpedraza1@yahoo.com.ar, oandrada@hotmail.com
njvaldez@c.exactas.unca.edu.ar, mteseira@gmail.com

Resumen

La lógica de la teoría de cambio se enfrenta con el problema de modelar el conocimiento dinámico, es decir, como el estado epistémico de un agente es modificado en un escenario dinámico. Usualmente, la nueva información a ser incorporada es asumida mediante una sentencia simple, como en el modelo *AGM*, pero la entrada epistémica también podría presentarse a través de un conjunto de sentencias. Por ello, es que resulta de interés estudiar las operaciones de *cambio múltiple*. Esta línea de investigación tiene como objetivo el estudio y ampliación de la utilidad del marco *AGM* mediante la investigación de nuevos tipos de cambio de creencias múltiples. Se estudiarán los *operadores de revisión* en donde algunos de ellos procesan la información con el apoyo de una *explicación*. También, se establecerá procedimientos de aplicación de estas construcciones en *sistemas argumentativos* que brindarán soporte a las capacidades de razonamiento de agentes que proveen información que puede ser incompleta y/o contradicto-

ria y de esta manera mejorar los mecanismos de toma de decisión automática. En tal sentido, se espera que los resultados obtenidos brinden una nueva perspectiva para desarrollar herramientas tecnológicas que funcionen desde un enfoque multi-agente.

Palabras Claves Cambio de Creencias Múltiples, Razonamiento Argumentativo, Programación Lógica Rebatible.

1. Contexto

Esta línea de investigación se realizará dentro del ámbito del Laboratorio de Investigación del Departamento de Ciencias de la Computación de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Catamarca. Esta asociado con el proyecto de investigación Bianual: “*Cambio de Creencias Múltiples en Sistemas Argumentativos Aplicado para Programación Lógica Rebatible*”. Financiado por el programa de desarrollo científico y tecnológico de la Secretaría de Ciencia y Tecnología: Consejo de Investigación, de la Universidad Na-

cional de Catamarca. Periodo: 01/01/2016 al 31/12/2017.

2. Introducción

Un sistema de revisión de creencias constituye un marco lógico para la modelización de la dinámica de conocimiento. Es decir, cómo modificamos nuestras creencias cuando recibimos nueva información. Usualmente, la nueva información a ser incorporada es asumida mediante una sentencia simple, como en la teoría AGM clásica [1, 2], pero la entrada epistémica también podría presentarse a través de un conjunto de sentencias. En este contexto, existen diversos tipos de operadores de cambios, como los operadores de *cambio priorizado*, los operadores *no priorizados*, los operadores *merging* u operadores de *cambio simétrico*, los de *cambios selectivos*, etc.

Debido a que las operaciones de cambios se llevan a cabo sobre el estado epistémico del agente, es que existen diferentes construcciones a ser usadas como modelos del estado epistémico, dos de los modelos más comúnmente usados son sobre *conjuntos de creencias* y sobre *bases de creencias*. Los conjuntos de creencias son conjuntos de sentencias lógicamente clausurados y las bases de creencias son conjuntos de creencias las cuales no son necesariamente clausuradas. Nuestra propuesta está centrada en las bases de creencias al tener como característica que son conjunto de sentencias no necesariamente cerradas, ello permiten que los cambios sobre el estado de las creencias representadas sean posibles desde el punto de vista computacional.

Dentro de este campo de la dinámica de conocimiento existen diversos tipos de operadores de revisión en donde algunos de ellos procesan la información con el apoyo de una *explicación*. Una *explicación* contiene un *ex-*

planans (las creencias que apoyan una conclusión) y un *explanandum* (la conclusión final). Cada explicación es un conjunto de sentencias con algunas restricciones. El rol de las explicaciones en la representación del conocimiento ha sido ampliamente estudiado en [7], siendo la misma la de racionalizar los hechos.

La revisión de creencias ha sido anexada a los *sistemas argumentativos* [6, 8], allí se define un nuevo modelo epistémico para que resulte posible la aplicación de operadores de revisión mediante explicaciones. En los *sistemas argumentativos* se utilizan creencias irrefutables y creencias refutables, y sus conclusiones son tentativas por lo que es necesario contar con un mecanismo de selección o criterio de preferencia entre argumentos [9, 12]. El concepto más general de razonamiento refutable se introdujo en Ciencias de la Computación como parte de la Inteligencia Artificial a mediados de los 80' por el filósofo John Pollock [11], abriendo el campo para alternativas más generales. Investigaciones en *Razonamiento No-Monótono*, *Programación Lógica* y *Argumentación* han dado resultados importantes, posibilitando desarrollar herramientas más poderosas para la Representación de Conocimiento y Razonamiento de Sentido Común. Los avances en esas áreas están logrando resultados útiles para otras áreas como el desarrollo de *Agentes Inteligentes* y aplicaciones de *Sistemas Multi-agente*. Los sistemas de argumentación resultan plausibles para la toma de decisiones, tales acciones pueden modelarse de una forma dialéctica usando argumentos a favor y en contra de cada opción [3, 4, 5].

Una aproximación a la implementación de este tipo de sistemas está dada por los servicios de razonamiento basados en *Programación Lógica Refutable*. La *Programación en Lógica Refutable* (Defeasible Logic Programming o DeLP) [9, 10] es uno de tales

formalismos, que combina resultados de la teoría de *Argumentación Rebatible y Programación Lógica*. DeLP provee un entorno apropiado para la construcción de aplicaciones que resuelvan problemas del mundo real asociados a información incompleta y contradictoria en dominios dinámicos.

3. Línea de Investigación y Desarrollo

Esta línea de investigación toma como punto de partida los aportes en la temática de los operadores de contracción de la teoría de cambio de creencias bajo clausulas Horn que han sido ampliamente investigado por los autores de este artículo [13, 14, 15, 16]. Uno de los objetivos principales es el estudio de las operaciones de cambio, en particular aquellos cambios de creencias que son inducidos a través de un conjunto de sentencias. En este contexto, consideramos dos tipos de cambios múltiples: el primero, los operadores de cambio múltiple priorizado en donde todas las nuevas creencias se supone que deben ser aceptadas, y el segundo, se refiere a operadores de mezcla (merging) que permite que creencias antiguas y nuevas jueguen roles simétricos dentro de un proceso de cambio. Para ambos tipos de cambios se analizan dos operadores, uno basada en kernels y la otra basada en conjuntos de restos.

Así también, dentro de la teoría de cambio, existen diversos tipos de operadores de revisión en donde algunos de ellos procesan la información con el apoyo de una explicación. Un operador de revisión puede modificar ya sea el conocimiento rebatible o no rebatible. El problema principal es determinar si alguna pieza de información es no rebatible o rebatible. Una solución simple podría ser la incorporación de conocimiento directamente en el conocimiento rebatible. Pero esta solución es demasiado simple y no es

muy realista. La calificación del conocimiento es dinámica, es decir, que evoluciona con el tiempo y en la incorporación de nueva información. Cuando un agente incorpora conocimiento por lo general incorpora conocimiento no rebatible. Es por ello, que en los sistemas argumentativos resulta necesario contar con criterios de preferencia entre argumentos. La implementación de este tipo de sistemas está dada por los servicios de razonamiento basados en Programación Lógica Rebatible.

Las Teorías de Cambio de Creencias y Argumentación (principales ejes de esta investigación) pertenecen al amplio campo de la Representación del Conocimiento, sin embargo, sus puntos focales básicos son diferentes. Ahora, las interrelaciones entre los dos campos están todavía en su mayor parte sin explorar. Tanto los campos de la teoría de la argumentación como la de revisión de creencias son de importancia sustancial para los sistemas multi-agente que se enfrentan a un uso intensivo en aplicaciones prácticas industriales. Sumado a ello, el creciente uso de métodos y herramientas de la teoría de cambio de creencias en la teoría de la argumentación y viceversa.

El presente proyecto, tiene por finalidad volcar resultados y mejoras que se obtengan en el desarrollo de modelos de revisión de creencias con sentencias múltiples combinado con argumentación y de esta manera agilizar las capacidades de razonamiento mediante la implementación de la programación lógica rebatible (DeLP).

4. Resultados y Objetivos

Diversas contribuciones relacionado a la temática de esta investigación, fueron presentados en el ámbito de las ciencias básicas, provocando un impacto directo en el desa-

rollo de áreas tecnológicas de vanguardia en Ciencias de la Computación, tales como la especificación de instituciones virtuales, desarrollo y automatización de mercados, subastas electrónicas a través de Internet, mejoras de mecanismos de toma de decisión automática, etc.

En esta línea de investigación se espera obtener los siguientes resultados:

- Ampliar la utilidad de las construcciones adaptadas con sentencias múltiples de tipo priorizados y simétricos a otros tipos de cambios y su aplicación con modelos de operadores de cambio (Partial meet, Kernel, Epistemic Entrenchment, entre otros).
- Investigar y determinar las interrelaciones entre dos campos de la representación de conocimiento como son la teoría de argumentación y cambio de creencias, considerando que son diferentes sus puntos focales, lo que promoverá el empleo de métodos basados en argumentación caracterizados por la adopción de elementos de la teoría de cambio de creencias.
- Formular procedimientos basados en programación lógica rebatible (DeLP), para mejorar las capacidades de razonamiento de agentes en operaciones de revisión ejecutados por sistemas argumentativos.

Respecto a los objetivos de esta investigación, se espera obtener contribuciones en el área de las ciencias básicas y en el ámbito de aplicaciones tecnológicas.

Por último, se integrará programas de capacitación e intercambio para los integrantes del proyecto y otras áreas afines. Estas actividades serán de carácter:

- técnica: estudio y manejo de herramientas informáticas como DeLP y

demás programas existentes en el campo de estudio; y

- metodológica: estrategias de investigación científica, mecanismos de publicación de artículos científicos, planificación de tutorías en ambientes presenciales y virtuales.

5. Formación de Recursos Humanos

Dentro de la temática de esta línea de investigación: *los operadores de contracción de la teoría de cambio de creencias bajo cláusulas Horn* han sido ampliamente investigado por parte de los autores de esta contribución, siendo el tema de tesis para alcanzar el grado de Magister en Ciencias de la Computación en la UNS (Argentina) de Néstor Jorge Valdez (uno de los autores de este artículo).

Así también, se espera que durante el desarrollo del proyecto, los integrantes del mismo que estén en proceso de elaboración de su tesis de posgrado puedan consolidar su formación en investigación, y que el trabajo realizado contribuya a su graduación.

Referencias

- [1] Alchourrón, Gärdenfors, and Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50, 1985.
- [2] Alchourrón and Makinson. On the logic of theory change: Contraction functions and their associated revision functions. *The Journal of Symbolic Logic*, 48:14–37, 1982.
- [3] Capobianco, Chesnevar, and Simari. Argumentation and the dynamics of

- warranted beliefs in changing environments. intl. *Journal on Autonomous Agents and Multiagent Systems (JAAMAS) 11 (Sept. 2005)*, pages 127–151, 2005.
- [4] Chesnevar, Maguitman, and Loui. Logical models of argument. *ACM Computing Surveys 32, 4 (Dec. 2000)*, pages 337–383, 2000.
- [5] Chesnevar, Maguitman, and Simari. Argument-based critics and recommenders: A qualitative perspective on user support systems. *Data and Knowledge Engineering 59, 2 (Nov. 2006)*, pages 293–319, 2006.
- [6] Falappa, García, and Simari. Belief dynamics and defeasible argumentation in rational agents. *In Proceedings of NMR 2004*, pages 164–170, 2004.
- [7] Falappa, Kern-Isberner, and Simari. Belief revision, explanations and defeasible reasoning. *Artificial Intelligence Journal*, 141:1–28, 2002.
- [8] Falappa, Kern-Isberner, and Simari. Argumentation in artificial intelligence. *In I. Rahwan, G. R. Simari (Eds.), Belief revision and argumentation theory*, pages 341–360, 2009.
- [9] García and Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4:95–138, 2004.
- [10] LIDIA. Delp web client and visualization tool. http://lidia.cs.uns.edu.ar/delp_client, pages–, 2007.
- [11] Pollock. Defeasible reasoning. *Cognitive science*, 11:481–518, 1987.
- [12] Prakken and Vreeswijk. Logical systems for defeasible argumentation. *In Handbook of Phil. Logic, D. Gabbay and F. Guenther, Eds. Kluwer*, pages 219–318, 2002.
- [13] Valdez and Falappa. Dinámica de conocimiento: Contracción múltiple en lenguajes horn. *XIX Congreso Argentino de Ciencias de la Computación, XIV Workshop Agentes y Sistemas Inteligentes (WASI), CACiC'2013*, pages–, 2013.
- [14] Valdez and Falappa. Dinámica de conocimiento: Contracciones horn a partir de ordenamientos epistémicos. *42JAIIO Jornadas Argentinas de Informáticas, ASAI 2013, 42 JAIIO'2013*, pages 206–209, 2013.
- [15] Valdez and Falappa. Implementación para bases de creencias horn de operadores de contracción múltiple. *XX Congreso Argentino de Ciencias de la Computación, XV Workshop Agentes y Sistemas Inteligentes (WASI), CACiC'2014*, 2014.
- [16] Valdez, Lara, Pedraza, and Teseira. Dinámica de conocimiento: Cambio de creencias múltiples. temática de investigación de inteligencia artificial en las ciencias de la computación. *JUCEN'15, Facultad de Ciencias Exactas y Naturales, UNCa. Catamarca 9 y 10 de diciembre de 2015 - Argentina*, pages 30–36, 2015.

Técnicas de razonamiento bajo incertidumbre aplicadas a la estimación de los destinos de viajes de ruta simple

Romina Stickar, Rodrigo René Cura

LINVI – Dpto. Informática – Facultad de Ingeniería
Universidad Nacional de la Patagonia San Juan Bosco
Puerto Madryn
Bvard. Brown 3051, (280) 4883499 int. 116
romistickar@gmail.com, rodrigo.renecura@gmail.com

Resumen

El presente artículo forma parte de la investigación necesaria para dar respuesta a la problemática del transporte público en la ciudad de Puerto Madryn. Uno de los objetivos del proyecto es hallar un modelo para estimar los puntos de descenso de una ruta única utilizando técnicas de razonamiento bajo incertidumbre. En particular, aquí presentamos una apreciación de algunas técnicas de representación de incertidumbre y la aplicabilidad de estas para la construcción de nuestro modelo.

Palabras clave: Simulación basada en agentes, Transporte, Sistemas de Soporte a la Decisión, Razonamiento bajo incertidumbre.

Contexto

Esta investigación se lleva a cabo dentro del proyecto “Desarrollo de una plataforma de software de simulación social para movilidad urbana en Puerto Madryn”, en el marco del programa

“Universidad y Transporte Argentino”. El mismo cuenta con la participación de alumnos y docentes de las Facultades de Ingeniería y de Humanidades y Ciencias Sociales de la Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB). Sedes Puerto Madryn y Trelew, en colaboración con el Centro Nacional Patagónico (CENPAT-CONICET).

El financiamiento del proyecto es realizado a través de la Secretaría de Políticas Universitarias (SPU) Ministerio de Educación.

Además, se espera que los resultados obtenidos sean un aporte para el proyecto de investigación “Clasificación de Información en Big Data mediante la utilización de técnicas de Inteligencia Artificial y Análisis de Redes Sociales” que el Laboratorio de Investigación en Informática (LINVI) acredita en la universidad.

Introducción

La movilidad urbana es un problema fundamental en los sectores de la población que viven en las áreas más alejadas de Puerto Madryn, en el acceso a las centralidades urbanas en general, y a

la universidad pública en particular. Para dar una posible solución a esta problemática se busca, como un objetivo general del proyecto, generar una plataforma de simulación social que permita dar apoyo a las decisiones de política pública relacionadas al transporte urbano [1]. Para poder generar esta plataforma es necesaria la construcción de un modelo de estimación de destinos de viajes unitarios, el cual se enfoca en estimar el *punto de descenso* para cada *punto de abordaje* de cada uno de los usuarios del sistema de transporte, es decir, dado un agente que entra en una ruta por el punto de ascenso, queremos determinar la parada en la cual descendió. Se puede pensar que cada una de las posibles paradas de descenso tiene una “fuerza de atracción” que aumenta la cantidad de pasajeros que bajarían allí, a su vez, la demanda del transporte tiene lugar en relación al *espacio*, dada por la distribución de las actividades en el mismo [2]. Entonces, se pueden considerar dichas actividades como puntos atractores de viajes en los momentos del día en que la actividad se encuentra activa.

Dado que solo nos concentraremos en viajes de ruta simple, es decir, sin transbordos, la representación de la fuerza de atracción será dependiente del área de influencia de la línea a la que pertenece la parada y de las actividades allí contenidas, en lugar de hacer foco en la red de transporte completa.

Nuestro enfoque hacia el problema nos lleva a representar los posibles viajes como un grafo dirigido valuado, donde los nodos representan paradas y el peso de los arcos la “intensidad” de los viajes entre una parada y otra. En el contexto de una simulación basada en agentes, dicha intensidad puede ser expresada con una técnica de representación de incertidumbre [3] en la generación de

intenciones de viaje para los agentes. Consideramos a continuación alguna de ellas.

La forma más intuitiva de representar lo anterior sería en base a frecuencias relativas para encontrar la distribución de probabilidad entre los distintos viajes posibles, esto es, la cantidad de viajes de la parada A a la parada B sobre la cantidad de viajes totales. Lamentablemente, el conjunto de datos disponibles brinda información limitada a nivel de viaje, debido a que solo hace referencia a los orígenes o destinos de los viajes, por ejemplo, desconocemos cuantos viajes se realizaron de A a B, pero sabemos cuántos pasajeros subieron en la parada A y cuantos descendieron en la parada B.

Otra forma de representación de la incertidumbre se enmarca en la teoría de evidencia desarrollada por Dempster y Shafer [4,5] que establece el concepto denominado función de creencia (belief function). Esto es interesante debido a que en la aplicación real la información que alimenta el modelo se presentará de manera progresiva y el mismo deberá ser capaz de absorberla manteniendo la coherencia, y es central a esta teoría la agregación de información.

Por último, surge considerar la técnica de intervalos de probabilidades, que en lugar de otorgar una probabilidad puntual trata de fijar límites razonables que contienen a la probabilidad buscada [6,7].

Líneas de Investigación, Desarrollo e Innovación

Dentro de lo transdisciplinario del proyecto, las líneas de investigación en cuanto a la computación están orientadas al análisis y comparación de las distintas

técnicas de razonamiento bajo incertidumbre, al modelado general de la plataforma, representación de datos recolectados, validación y calibración de la simulación.

Resultados y Objetivos

Actualmente el proyecto se encuentra en estado avanzado. Se ha recopilado información de las distintas líneas de colectivo que hay en la ciudad, se han realizado encuestas entre distintas personas que concurren a los diferentes ámbitos educativos. Además se hizo el recorrido de unas de las líneas de colectivo y se observó la cantidad de gente que ascendía y descendía en cada una de las paradas.

A partir de esta recolección de información, pudimos evaluar a fondo la aplicación de las técnicas de razonamiento antes mencionadas. Como adelantábamos la frecuencia relativa no pudo ser aplicada por la naturaleza de la información. La estrategia de funciones de creencia que en principio parecía más prometedora finalmente no resultó ser apta para la tarea debido a la información con la que contamos. Finalmente, el marco teórico de la técnica de intervalo de probabilidades nos resulta adecuado y se está evaluando su aplicación para nuestro conjunto de datos.

Los objetivos de la investigación se pueden resumir en:

- En función de la ruta a analizar y sus paradas, determinar las áreas de influencia de las mismas.
- Estudiar y analizar los métodos de razonamiento bajo incertidumbre para poder asignar un valor a cada una de las paradas.

- Realizar una evaluación comparativa para analizar la predictibilidad de los métodos estudiados.

Como culminación del proyecto se pretende haber desarrollado una plataforma de software para la simulación social que pueda ser aplicada en el análisis de accesibilidad hacia cualquier destino urbano y que permita ayudar a la generación de planes de acción en la planificación del transporte local.

Formación de Recursos Humanos

El proyecto permite la formación de estudiantes de grado y posgrado de las disciplinas de Ciencias de la Computación y las Ciencias Humanas y Sociales. El equipo de trabajo está formado por investigadores y becarios CENPAT-CONICET y alumnos y docentes de la Facultad de Ingeniería y Humanidades y Ciencias Sociales de la UNPSJB.

La dirección del proyecto está a cargo del Dr. Julio Vezub y co-dirigido por la Lic. Romina Sticker.

El Lic. Rodrigo René Cura y el Lic. Sergio Kaminker participan del proyecto en el marco de una beca interna doctoral de CONICET, para los doctorados en Ciencias de la Computación y Sociología respectivamente.

Los alumnos de grado Roberto Voogt y Roxana Velasquez reciben una beca, desde los fondos del proyecto, como incentivo para la finalización de su carrera de Licenciatura en Informática y Trabajo Social.

Referencias

1. René Cura, R. & Sticker, R. & Buckle, C. & Voogt, R. & Kaminker, S. & Velazquez, R. & Vezub, J. (2015) *Desarrollo de una plataforma de software de simulación social para movilidad urbana en Puerto Madryn*.
2. De Dios Ortúzar, J., & Willumsen, L. G. (2008). *Modelos de transporte* (Vol. 1) Ed. Universidad de Cantabria.
3. Halpern, J. Y. (2003). *Reasoning about uncertainty* (Vol. 21). Cambridge: MIT press.
4. Shafer, G. (1976). *A mathematical theory of evidence*. Princeton: Princeton university press.
5. Shafer, G. (2015). Dempster's rule of combination. *International Journal of Approximate Reasoning*.
6. Kyburg Jr, H. E. (1998). Interval-valued probabilities. *Imprecise Probabilities Project*.
7. Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24(2), 125-148.

Técnicas de Inteligencia Artificial aplicadas a problemas científico-tecnológicos

Javier Izetta Riera, Nilda M. Pérez Otero
GIDIA / Facultad de Ingeniería / Universidad Nacional de Jujuy
Ítalo Palanca 10, +54 (388) 4221587
{javierizetta, nilperez}@gmail.com

Resumen

Cada vez es más intensivo el uso de técnicas de inteligencia artificial en la mayoría de las actividades humanas y en particular en el área científico-tecnológico. Algunas de las áreas, que muestran tanto la diversidad de campos a los que pueden aplicarse como la utilidad que pueden alcanzar, son: Bioinformática, Robótica, Tecnología de los alimentos y Recuperación de información.

Una de las tareas más requeridas es la generación automática de modelos que resuman las características que debe cumplir un elemento para pertenecer a una determinada categoría, tarea que se conoce como clasificación automática. Los métodos computacionales desarrollados para tal fin forman parte de lo que se conoce como Aprendizaje Automatizado (Machine Learning). Otra tarea importante es la optimización de funciones, en las últimas décadas diferentes tipos de algoritmos de optimización han sido desarrollados para resolver una gran cantidad de problemas. El principal desafío radica en problemas que presentan una función objetivo altamente no lineal y no convexa, esto dificulta garantizar la localización del mínimo global. Por lo tanto, la necesidad de encontrar nuevas técnicas que proporcionen un mejor desempeño en este tipo de problemas sigue aún vigente y

hacen de esto un campo excitante para trabajar.

Palabras clave: Metaheurísticas, Aprendizaje Automatizado, Inteligencia Artificial.

Contexto

La línea de investigación aquí presentada se encuentra inserta en el proyecto *Técnicas de Inteligencia Artificial aplicadas a problemas científico-tecnológicos*, ejecutado a partir de 2016 por el Grupo de Investigación y Desarrollo en Informática Aplicada (GIDIA) de la Facultad de Ingeniería de la Universidad Nacional de Jujuy.

El proyecto se encuentra acreditado y financiado por la Secretaría de Ciencia y Técnica y Estudios Regionales de la Universidad Nacional de Jujuy y se encuentra bajo el Programa de Incentivos.

Introducción

La Inteligencia Artificial (AI) propone nuevas herramientas computacionales que brindaron solución a muchos problemas complejos del mundo real. Estas herramientas se utilizan en una variedad de aplicaciones que van desde el modelado, la clasificación, el reconocimiento de patrones hasta el análisis de datos. Las técnicas de inteligencia artificial se vienen utilizando

en la ciencia y la tecnología de los alimentos para la clasificación, el modelado y la optimización de procesos, el control de calidad de los alimentos, la predicción de propiedades reológicas de masa, la clasificación de vinos en función del contenido de antocianinas, entre otros [1]. Asimismo en el área de la ingeniería industrial se comenzaron a incorporar herramientas de inteligencia artificial, a tal punto que desde en el año 2015 se realizó la Conferencia Internacional de Inteligencia Artificial e Ingeniería Industrial (AEII2015) [2].

Particularmente, una rama de la inteligencia artificial que en los últimos años ha tenido gran crecimiento es el Aprendizaje Automatizado (Machine Learning) la cual propone nuevas herramientas computacionales que sirven de apoyo para tareas de investigación en distintas áreas.

Aprendizaje Automatizado

En las últimas décadas el incremento de métodos automáticos de medición y almacenamiento de datos produjo una revolución en las ciencias, donde la cantidad de nuevas observaciones disponibles supera ampliamente nuestra capacidad actual de modelado. Esta situación produjo la necesidad de desarrollar programas de computación que, a partir de un conjunto de datos, generen un modelo que produzca salidas aproximadas a las del proceso que está implícito en los datos observados. El aprendizaje automatizado intenta comprender los mecanismos por los cuales se adquiere el conocimiento a partir de los datos/experiencia y con ello lograr automatizar la etapa de modelado [3]. El conjunto de datos de entrada de un programa de aprendizaje automatizado consiste en descripciones de objetos de un universo acompañado con un valor de salida asociado. Cada objeto es

considerado un ejemplo del universo que se desea conceptualizar. Generalmente estos ejemplos son tuplas de la forma (atributos; salida). El proceso de aprendizaje consiste entonces en la búsqueda dentro de un espacio H , de todas las hipótesis, la hipótesis f que mejor aproxime la función original F que dio origen al conjunto datos de entrada.

Clasificación

Los problemas de clasificación son una parte del aprendizaje automatizado [4] en los que el objetivo es asignar a los datos $D = \{x_1; \dots ; x_n\}$ un número discreto de valores $c \in C = \{c_1; \dots ; c_n\}$, también llamados clases o categorías. Cada objeto tiene asignada una clase o etiqueta real $l(x_i)$ conocida. Para un problema con c clases $l(x_i)$ puede tomar c valores discretos distintos. El objetivo entonces es encontrar una función clasificadora f tal que para cada objeto x_i sea $f(x_i) = l(x_i)$, donde $f(x_i)$ es la etiqueta o clase asignada por f a un objeto.

Muchos métodos de clasificación han sido introducidos en los últimos años, algunos de estos son: Redes Neuronales Artificiales [5], Máquinas de Vectores Soporte [6] y Random Forest [7]. Los cuales fueron aplicados a una amplia variedad de problemas científicos tecnológicos [8], [9], [10], [11], [12] demostrando buenos resultados. Esto nos lleva a proponer la utilización de estos métodos para resolver diferentes problemas en áreas como Bioinformática, Robótica, Recuperación de información y particularmente en Tecnología de los alimentos.

Metaheurísticas

Las metaheurísticas, en su definición original, son métodos de resolución que orquestan una interacción entre los procesos de mejora local y estrategias de mayor nivel para crear un proceso capaz

de escapar de óptimos locales y realizando una búsqueda robusta en el espacio de soluciones. Con el tiempo, estos métodos han llegado a incluir cualquier procedimiento que emplee estrategias para superar la trampa de la optimalidad local en espacios de soluciones complejos, especialmente aquellos procedimientos que utilizan una o más estructuras locales como un medio para definir movimientos admisibles para la transición de una solución a otra, o para construir o destruir soluciones en procesos constructivos y destructivos [13].

Varias herramientas y mecanismos surgidos a partir de la creación de los métodos metaheurísticos han demostrado ser eficaces, tanto es así que las metaheurísticas se convirtieron en el centro de atención en los últimos años como el enfoque preferido para resolver problemas de optimización complejos, en particular problemas de naturaleza combinatoria. Si bien, las metaheurísticas no son capaces de garantizar la optimalidad de las soluciones que encuentran, los procedimientos exactos o métodos de convergencia local (que en teoría pueden brindar tal garantía, si se les permite correr el tiempo suficiente) a menudo son incapaces de encontrar soluciones cuya calidad sea similar a la obtenida por las principales metaheurísticas, en particular, en la resolución de problemas del mundo real donde se evidencia su eficiencia y eficacia para resolver problemas grandes y complejos. La aplicación de metaheurísticas comprende un gran número de áreas y campos disciplinarios, siendo algunos de ellos:

- Diseño de ingeniería, optimización de topologías y la optimización estructural en electrónica y

VLSI, aerodinámica, dinámica de fluidos, telecomunicaciones y robótica.

- Machine learning y minería de datos en bioinformática y biología computacional, y finanzas.
- Modelado de sistemas, simulación e identificación en química, física y biología; control, señal, y procesamiento de imágenes.
- Planificación de problemas de enrutamiento, problemas de planificación, programación y producción de robots, logística y transporte, gestión de la cadena de suministro, y otros.

En el diseño de una metaheurística se deben tener en cuenta dos criterios contradictorios: la exploración del espacio de búsqueda (diversificación) y la explotación de las mejores soluciones encontradas (intensificación) [14]. Las regiones más promisorias están determinadas por la “buenas” soluciones encontradas. En la intensificación, estas regiones se exploran más a fondo con la esperanza de encontrar mejores soluciones. En la diversificación, las regiones no exploradas deben ser visitadas para asegurarse de que todas las regiones del espacio de búsqueda se exploran de manera uniforme y que la búsqueda no se limita a sólo un número reducido de regiones.

Las metaheurísticas pueden clasificarse desde varios puntos de vista, [14] presenta las siguientes alternativas de clasificación:

- ***Inspiradas o no inspiradas en la naturaleza:*** varias metaheurísticas se inspiran en procesos naturales, por ejemplo: algoritmos evolutivos y sistemas artificiales inmunes se basan en principios de la biología; colonias de hormigas, abejas, y optimización de enjambre de partículas están inspirados en la

inteligencia colectiva de diferentes especies (ciencias sociales); y el recocido simulado utiliza principios de la física.

- **Métodos con o sin uso de memoria:** algunos algoritmos metaheurísticos no emplean elementos de memoria, es decir, no se utiliza información extraída de forma dinámica durante la búsqueda. Algunos algoritmos de esta clase son la búsqueda local, GRASP, y recocido simulado. Otras metaheurísticas usan una memoria que contiene parte de la información extraída en línea durante la búsqueda. Por ejemplo, las memorias a corto plazo y largo plazo en la búsqueda tabú.
- **Deterministas o estocásticas:** una metaheurística determinista resuelve un problema de optimización mediante la toma de decisiones deterministas (por ejemplo, búsqueda local, búsqueda tabú). Las metaheurísticas estocásticas aplican algunas reglas aleatorias durante la búsqueda (por ejemplo, recocido simulado, algoritmos evolutivos). En los algoritmos deterministas, la misma solución inicial dará lugar a la misma solución final, mientras que en las metaheurísticas estocásticas, diferentes soluciones finales se pueden obtener de la misma solución inicial. Esta es una característica a tener en cuenta en la evaluación del desempeño de los algoritmos.
- **Búsqueda poblacional o búsqueda de única solución:** los algoritmos basados en una sola solución (por ejemplo, recocido simulado) manipulan y transforman una única solución durante la búsqueda, mientras que en los algoritmos poblacionales (por ejemplo,

enjambre de partículas, algoritmos evolutivos) es toda una población de soluciones la que evoluciona. Estas dos clases tienen características complementarias: las metaheurísticas de una sola solución están orientadas a la explotación, tienen el poder de intensificar la búsqueda en las regiones locales. Las metaheurísticas basadas en población se orientan a la exploración, ya que permiten una mayor diversificación en el espacio de búsqueda conjunto.

- **Iterativas o avaras:** en los algoritmos iterativos, se parte de una solución completa (o población de soluciones) y se transforman en cada iteración utilizando algunos operadores de búsqueda. Los algoritmos golosos, parten de una solución vacía, y en cada paso se asigna una variable de decisión del problema hasta que se obtiene una solución completa. La mayoría de las metaheurísticas son algoritmos iterativos.

Líneas de Investigación, Desarrollo e Innovación

Esta línea de investigación consiste en analizar técnicas de metaheurísticas y de aprendizaje automatizado para aplicarlas al desarrollo de algoritmos robustos para problemas de optimización global y clasificación, involucrados en tareas científicas tecnológicas de distintas áreas.

Resultados y Objetivos

El proyecto, que se realizará en el bienio 2016-2017 tiene como objetivo general **objetivo general** desarrollar algoritmos que implementen técnicas de inteligencia artificial, una combinación de

las ya existentes u otras técnicas para resolver problemas científico-tecnológicos de interés. Entre los objetivos específicos se encuentran:

- Establecer conclusiones en cuanto desempeño y eficiencia de las técnicas de inteligencia artificial frente a las tradicionales.
- Proponer y desarrollar algoritmos para la resolución del conjunto de problemas planteados.
- Determinar desempeño numérico y eficiencia de los nuevos algoritmos.

Formación de Recursos Humanos

El equipo de trabajo está integrado por docentes-investigadores y alumnos de las Universidad Nacional de Jujuy. El equipo actual cuenta con 5 ingenieros en informática y 1 alumno. Se prevé la finalización de una tesis de doctorado, realización de una tesis de maestría y una tesina de grado.

Referencias

[1] Lee, C. J. (Editor). International Conference on Artificial Intelligence and Industrial Engineering 2015 (AIIE 2015). Proceedings of a meeting held 26-27 July 2015, Phuket, Thailand. ISBN 9781510806450. Atlantis Press.

[2] Sofu, B.A., Demir, N., Ekinçi, F.Y., 2007. Gıda Bilimi Ve Teknolojisi Alanında Yapay Zeka Uygulamaları (Applications of Artificial Intelligence in Food Science and Technology Area). GIDA (2007) 32 (2): 93-99.

[3] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.

[4] Mitchell, T. Machine Learning. McGraw-Hill series in computer science. McGraw-Hill, 1997.

[5] Hassoun, M. H. Fundamentals of artificial neural networks. MIT press, 1995.

[6] Vapnik, V. Statistical learning theory. 1998, 1998.

[7] Breiman, L. Random forests. *Machine learning*, 45 (1), 5-32, 2001.

[8] Pavlidis, P., Wapinski, I., & Noble, W. S. (2004). Support vector machine classification on the web. *Bioinformatics*, 20(4), 586-587.

[9] Byvatov, E., Fechner, U., Sadowski, J., & Schneider, G. (2003). Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *Journal of Chemical Information and Computer Sciences*, 43(6), 1882-1889.

[10] Gardner, G. G., Keating, D., Williamson, T. H., & Elliott, A. T. (1996). Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *British journal of Ophthalmology*, 80(11), 940-944.

[11] Lempitsky, V., Verhoek, M., Noble, J. A., & Blake, A. (2009). Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography. In *Functional Imaging and Modeling of the Heart* (pp. 447-456). Springer Berlin Heidelberg.

[12] Miller, K., Huettmann, F., Norcross, B., & Lorenz, M. (2014). Multivariate random forest models of estuarine-associated fish and invertebrate communities. *Mar Ecol Prog Ser*, 500, 159-174.

[13] Glover, F. W. & Kochenberger, G. A., editores (2003). *Handbook of Metaheuristics*, volume 114 of International Series in Operations Research & Management Science. Springer.

[14] Talbi, E.-G. (2009). *Metaheuristics: From Design to Implementation*. Wiley Publishing.

Algoritmos Evolutivos híbridos para el diseño y operación eficiente de una red de distribución de agua potable

Villagra A., Pereyra G., Molina D., Seron N., Goupillat C., Varas V., Montenegro C., Lasso M., Pandolfi D.

Laboratorio de Tecnologías Emergentes (LabTEM)

Instituto de Tecnología Aplicada - Unidad Académica Caleta Olivia

Universidad Nacional de la Patagonia Austral

{avillagra, gpereyra, dmolina, nseron}@uaco.unpa.edu.ar, goupinge@yahoo.com.ar,

{vvaras, cmontenegro, mlasso, dpandolfi}@uaco.unpa.edu.ar

Leguizamón G.

Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)

Departamento de Informática - Universidad Nacional de San Luis

legui@unsl.edu.ar

Resumen

Los Algoritmos Evolutivos (AEs) han sido ampliamente utilizados de forma exitosa en las últimas dos décadas para resolver diferentes problemas de redes de distribución de agua. El problema de diseño de una red de distribución de agua (RDA) ha sido reconocido como un problema NP-duro que no puede ser resuelto fácilmente usando técnicas matemáticas tradicionales. En esta línea de trabajo se utilizan dos algoritmos Evolutivos (Algoritmo Genético Celular – cGA- y Algoritmo CHC - *Crossover elitism population, Half uniform crossover combination, Cataclysm mutation*) no tan utilizados en la literatura para resolver este problema pero que han tenido resultados exitosos en distintos problemas de optimización. Además, se pretende incorporar versiones híbridas de ambos algoritmos actualmente utilizadas en diversos problemas de optimización en el marco otros proyectos de investigación llevados a cabo en el Laboratorio de Tecnologías Emergentes. Con estos algoritmos se intenta brindar soluciones al diseño

de una red eficiente de distribución de agua potable para la localidad de Caleta Olivia.

Palabras clave: Algoritmos Evolutivos, Red de agua potable, Optimización, cGA, CHC.

Contexto

La línea de investigación presentada en este documento se lleva a cabo en el Laboratorio de Tecnologías Emergentes (LabTEM) en el marco del Programa de Investigación en Ciencia y Tecnología de la Universidad Nacional de la Patagonia Austral, Unidad Académica Caleta Olivia. En el ámbito de una convocatoria PDTS-UNPA, proyecto denominado: “Metaheurísticas aplicadas al diseño y operación óptima de una red urbana de distribución de agua potable”.

Introducción

Actualmente, las redes de distribución de agua juegan un papel vital e importante en

la vida de la sociedad. Estas redes se componen de tuberías, tanques, bombas, depósitos, válvulas y algunos otros componentes. Se pueden considerar diferentes configuraciones de estos componentes en relación con algunas restricciones para proporcionar agua a los consumidores. El objetivo de un diseñador es minimizar el costo de la red y puede considerarse como un problema de optimización con diferentes aspectos tales como el costo de mantenimiento, diseño de la disposición, la fiabilidad, la selección del material, requerimientos de la demanda, entre otros.

Las técnicas de optimización tradicionales como la programación lineal [3, 6, 15], la programación dinámica [17] y la programación no lineal [8] han sido empleadas principalmente para resolver la minimización de costo del problema de diseño de la RDA. Sin embargo, estos modelos sólo se han aplicado a tamaño de problemas pequeños (simplificaciones) y suposiciones debido a la complicada naturaleza del problema. De hecho, el diseño de la RDA está clasificado como un problema NP-duro [22].

Los AEs se han aplicado a una pléthora de problemas a través de una amplia variedad de disciplinas. Su relativa simplicidad y su capacidad de trabajar eficientemente en nuevos problemas han hecho que sean adoptados en campos tan diversos como la ingeniería, la economía y la robótica. Como es de esperar, los AEs también se han aplicado a problemas de redes de distribución de agua con un alto grado de éxito, por ejemplo, en el campo de la recuperación de aguas subterráneas [14], el control de la morfología del lecho fluvial [13], la determinación de las características hidráulicas de producción pozos [10] y en particular en el campo de la optimización de la distribución de agua de la red [4, 16]. También diferentes metaheurísticas

se han aplicado al problema de RDA como por ejemplo, Algoritmos genéticos en [9, 11], Recocido Simulado [5, 19], Optimización basado en cúmulo de partículas [1] y Optimización basada en sistemas de hormigas [23]. Si bien existe una importante cantidad de publicaciones, el problema de diseño y operación eficiente de una RDA es un problema de frontera abierta es decir, que está abierto a encontrar mejores soluciones. Por esta razón, planteamos la utilización de dos potentes algoritmos cGA y CHC.

El cGA es un algoritmo basado en población descentralizada en el que las soluciones tentativas evolucionan en los vecindarios solapados [2, 18]. El algoritmo CHC [7, 20] es un Algoritmo Genético (AG) no tradicional que combina una estrategia de selección conservativa que siempre preserva los mejores individuos encontrados. Hemos obtenido resultados promisorios con ambos algoritmos y sus versiones híbridas aplicados a problemas de optimización [12, 21]. En esta línea de trabajo los utilizaremos para abordar diferentes problemas de la red de distribución de agua, tomando además como caso de estudio la red actual de la ciudad de Caleta Olivia, provincia de Santa Cruz, Argentina.

Líneas de investigación y desarrollo

Una red de distribución de agua es el conjunto de tuberías, accesorios y estructuras que conducen el agua desde embalses o cisternas de servicio hasta las tomas domiciliarias o hidrantes públicos. Su finalidad es proporcionar agua a los usuarios para consumo doméstico, público, comercial, industrial y para condiciones extraordinarias como extinguir incendios. La red debe proporcionar este servicio todo el

tiempo, en cantidad suficiente, con la calidad requerida y a una presión adecuada. Las conexiones entre todos los componentes de la red de distribución pueden variar, esto hace que la red sea compleja, en cuanto a su comportamiento y diseño. Siendo la distribución de agua un factor crítico, uno de los principales desafíos es encontrar la red que cumpla con determinadas restricciones de presión y caudal a un costo razonable. La optimización de una red de distribución está determinada por tres fases diferentes: (1) Topológica, el nivel de decisión es estratégico e involucra algunas variables de decisión como conexiones, válvulas y bombas. (2) Diseño de la red, cuyo nivel de decisión es táctico e involucra como variables de decisión los diámetros de las tuberías y la rugosidad. (3) Planificación que involucra el nivel de decisión operacional tomando como variables de decisión el control de válvulas y bombas.

Atendiendo a estas fases ésta línea de investigación como objetivos generales pretende a nivel estratégico ayudar a evaluar la topología de la red actual; a nivel táctico determinar el impacto del diámetro y rugosidad de las tuberías en cuanto al diseño. Además, colaborar en la planificación de las prioridades de distribución, la planificación y control de las válvulas.

Resultados obtenidos/esperados

En cuanto a los resultados de ésta línea de investigación y desarrollo se espera lograr:

- Reconstruir la topología de la red troncal de abastecimiento de agua de la ciudad de Caleta Olivia.
- Caracterizar las restricciones de abastecimiento de agua a la red.

- Caracterizar la demanda de agua en cada válvula de provisión de la red troncal.
- Determinar las presiones en los distintos puntos de la red troncal de distribución.
- Proponer el diseño óptimo de la red troncal a través de los algoritmos cGA, CHC y sus versiones híbridas.
- Simular la red de distribución propuesta para Caleta Olivia.
- Comparar los resultados propuestos con aplicaciones de distribución libre (tal como EPANet¹- programa que realiza simulaciones para la RDA).
- Simular la red actual con restricciones de reservorio con algoritmos CHC y cGA para programar la distribución óptima de agua.
- Analizar los resultados en conjunto con la empresa Servicios Públicos Sociedad del Estado (SPSE).

Con estos resultados se pretende a través del uso de Algoritmos Evolutivos desarrollar una herramienta capaz de generar soluciones a distintos escenarios de la planificación y operación de una red urbana de distribución de agua potable. Cabe resaltar que esta herramienta podrá ser en el futuro fácilmente adaptada para aplicarse en el diseño de otros tipos de redes tal como una red cloacal.

Actualmente, se está trabajando en la preparación de los datos de la red troncal de Caleta Olivia. Además, se está realizando el estudio de los algoritmos propuestos sintonizándolos a través de su aplicación a redes de distribución de agua utilizadas en la literatura (Red de Hanoi, Red de New York y Red de Alperovits y Shmir).

Formación de recursos humanos

¹ <http://www.instagua.upv.es/Epanet/>

Un integrante del proyecto está desarrollando su tesis de Doctorado orientada a esta línea de investigación.

Se cuenta con un becario de grado realizando su trabajo de fin de carrera en este tema.

Referencias

- [1] K. Aghdam, I. Mirzaee, N. Pourmahmood, M. Aghababa. Design of water distribution networks using accelerated momentum particle swarm optimisation technique. *Journal of Experimental & Theoretical Artificial Intelligence*, Vol. 26, No. 4, 459–475, 2014.
- [2] E. Alba and B. Dorronsoro. *Cellular Genetic Algorithms*. Springer, 2008.
- [3] E. Alperovits, U. Shamir, Design of optimal water distribution systems, *Water Resour. Res.* 13 (6) 885–900, 1977.
- [4] W. Bi, G. Dandy, H. Maier. Improved genetic algorithm optimization of water distribution system design by incorporating domain knowledge. *Environ. Modell. Softw* 69, 370-381, 2015 <http://dx.doi.org/10.1016/j.envsoft.2014.09.010>
- [5] M. Cunha, J. Sousa. Water distribution network design optimization: simulated annealing approach. *J Water Resour Plan Manage*, 125(4):215–21, 1999.
- [6] G. Eiger, U. Shamir, A. Ben-Tal, Optimal design of water distribution networks, *Water Resour. Res.* 30 (9), 2637–2646, 1994.
- [7] L. Eshelman. The CHC Adaptive Search Algorithm: How to Have Safe Search When Engaging in Nontraditional Genetic Recombination. In *Foundations of Genetic Algorithms*, pages 265- 283. Morgan Kaufmann, 1991.
- [8] O. Fujiwara, D. Khang, A two-phase decomposition method for optimal design of looped water distribution networks, *Water Resour. Res.* 26 (4), 539–549, 1990.
- [9] D. Goldberg, C. Kuo. Genetic algorithms in pipeline optimization. *J Comput Civil Eng*, 1(2):128–41, 1987.
- [10] M. Jha, G. Nanda, M. Samuel. Determining hydraulic characteristics of production wells using genetic algorithm. *Water Resour. Manag.* 18 (4), 353-377, 2004.
- [11] M. Kadu, R. Gupta, P. Bhawe. Optimal design of water networks using a modified genetic algorithm with reduction in search space. *J Water Resour Plan Manage*, 134(2):147–60, 2008.
- [12] D. Molina, D. Pandolfi, A. Villagra, y G. Leguizamón. Applying CHC Algorithms on Radio Network Design for Wireless Communication. *Computer Science & Technology Series, XX Argentine Congress of Computer Science*, Selected papers, pag 27-37, 2015.
- [13] J. Nicklow, O. Ozkurt, Jr. Bringer. Control of channel bed morphology in large-scale river networks using a genetic algorithm. *Water Resour. Manag.* 17 (2), 113-132, 2003.
- [14] Piscopo, J. Kasprzyk, R. Neupauer. An iterative approach to multiobjective engineering design: optimization of engineered injection and extraction for enhanced groundwater remediation. *Environ. Model. Softw*, 2014.
- [15] G. Quindry, E. Brill, J. Liebman, Optimization of looped water distribution

systems, *J. Environ. Eng.* 107 (4), 665–679, 1981.

[16] D. Savic, G. Walters. Genetic algorithms for least-cost design of water distribution networks. *J. Water Resour. Plann. Manag.* 123 (2), 67-77, 1997

[17] J.C. Schaake, D. Lai, Linear programming and dynamic programming applications to water distribution network design, Rep. No. 116, Hydrodynamics Laboratory, MIT, Cambridge, MA, 1969.

[18] M. Tomassimi. The parallel genetic cellular automata: Application to global function optimization. In: Albrecht R, Reeves C, Steele N (eds) *International Conference on Artificial Neural Networks and Genetic Algorithms*, Springer-Verlag, Heidelberg, pp 385–391, 1993.

[19] J. Tospornsampan, I. Kita , M. Ishii, Y. Kitamura. Split-pipe design of wáter distribution network using simulated annealing. *Int J Comput Inform Syst Sci Eng*, 1(3):153–63, 2007.

[20] L. Vélez. Enfoques no estándar de algoritmos evolutivos en un dilema de optimización. *Mutis*, 2(2):126-138, 2012.

[21] A. Villagra, G. Leguizamón, and E. Alba. Active components of metaheuristic in cellular genetic algorithms. *Soft Computing*, pages 1-15, 2014.

[22] D. Yates, A. Templemen, T. Boffey, The computational complexity of the problem of determining least capital cost designs for water supply networks, *Eng. Optim*, 7 (2) 142–155, 1984.

[23] A. Zecchin, H. Maier, A. Simpson, M. Leonard, J. Nixon. Ant colony optimization applied to water distribution system

design: Comparative study of five algorithms. *Journal of Water Resources Planning and Management*, 133 (1):87-92, 2007

Algoritmo CHC aplicado a la optimización de coberturas de señales de radio frecuencia en comunicaciones inalámbricas en locaciones petroleras

Molina D., Villagra A., Villagra S., Valdez J.C., Rasjido J., Mercado V., Pandolfi D.

Laboratorio de Tecnologías Emergentes (LabTEM)

Instituto de Tecnología Aplicada (Caleta Olivia)

Universidad Nacional de la Patagonia Austral

{dmolina, avillagra, svillagra, jcvldez, jrasjido, vmercado, dpandolfi}@uaco.unpa.edu.ar

Leguizamón G.

Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)

Departamento de Informática - Universidad Nacional de San Luis

legui@unsl.edu.ar

Resumen

La difusión de los servicios de comunicación inalámbricos (teléfono, internet, etc.) está creciendo continuamente en estos días. Desafortunadamente, el costo de los equipos para proporcionar el servicio con la calidad adecuada es alta. Por lo tanto, la selección de un conjunto de puntos geográficos que permiten la cobertura óptima de una señal de radiofrecuencia, reduciendo al mínimo el uso de los recursos es esencial. A esta tarea se la denomina diseño de red de radio (siglas RND del *inglés Radio Network Design*) y es un problema NP-duro, por lo que el uso de las metaheurísticas es un enfoque viable para su resolución. Las metaheurísticas son métodos que integran procedimientos de mejora local y estrategias de alto nivel para realizar una búsqueda robusta en el espacio del problema. Enfoques avanzados del algoritmo CHC (*Cross generational elitist selection Heterogeneous recombination Cataclysmic mutation algorithm*) son aplicados a la optimización de coberturas eficientes de señales de radio frecuencia en

comunicaciones inalámbricas en locaciones petroleras.

Palabras clave: Algoritmos Evolutivos, Optimización de Servicios Inalámbricos, Red de Radio Frecuencia, CHC.

Contexto

La línea de investigación presentada en este documento se lleva a cabo en el Laboratorio de Tecnologías Emergentes (LabTEM) en el marco del Programa de Investigación en Ciencia y Tecnología de la Universidad Nacional de la Patagonia Austral, Unidad Académica Caleta Olivia. En el ámbito de una convocatoria de proyectos de I&D UNPA, proyecto denominado: “Metaheurísticas avanzadas aplicadas al diseño eficiente de redes de radio frecuencia en comunicaciones inalámbricas en locaciones petroleras”.

Introducción

El proceso de producción de petróleo (extracción, transporte y venta), involucra una amplia serie de controles en cada una de

sus etapas. Por lo tanto, el monitoreo y control del proceso es fundamental para cualquier compañía petrolera. Los sistemas de supervisión, control y adquisición de datos (SCADA) permiten la gestión y control de cualquier sistema local o remoto gracias a una interfaz gráfica, *Human Machine Interface* (HMI), que comunica al usuario con el sistema. Para que el SCADA funcione se debe alimentar el sistema con señales de sensores ubicados a kilómetros de distancia, permitiendo controlar de forma automática los posibles puntos de falla que pueden provocar derrames generando pérdidas y un impacto ambiental. Para lograr la comunicación entre tanta cantidad de dispositivos (miles de pozos petroleros por yacimiento) se necesita una red de radio con una cobertura eficiente. Básicamente la red de radio es un conjunto de dispositivos interconectados de manera inalámbrica compuesto por un emisor y un receptor. El vínculo entre ambos puntos se realiza utilizando radio frecuencia basada en ondas hertzianas y el medio que transporta estas ondas es el aire. Las ondas al desplazarse por el espacio libre están sujetas a ruidos (atenuación, reflexión, refracción y difracción) provocados por fenómenos atmosféricos y obstáculos. Como consecuencia se generan pérdidas haciendo que muchas veces el receptor no sea capaz de decodificar el mensaje. Según Penin, [17], el problema del posicionamiento de antenas puede ser descrito como: dado un conjunto de sitios candidatos, un área geográfica discretizada y un conjunto de puntos que deben comunicarse entre sí, se debe seleccionar del conjunto de sitios candidatos un subconjunto de sitios, que maximicen la cobertura utilizando el mínimo de recursos, cumpliendo con la estimación de tráfico y el umbral de recepción de señal entre los puntos. Del diseño de la red de radio surgen dos aspectos: a) Topográfico: involucra todo lo relacionado con la posición

geográfica de las estaciones y la topología del salto, es decir, lo referido a las cotas y los accidentes del terreno, las coordenadas y azimut de las estaciones y la distancia entre estaciones. Para representar el área que se desea cubrir se necesita un método que permita generar un mapa digital con los valores más representativos del terreno y que sea la referencia de obstáculos para el modelo de propagación seleccionado. Se desarrollará una representación del terreno sobre el cual se realizará la distribución de las radio bases (antenas). b) Radio eléctrico: abarca todo lo vinculado a la propagación de la onda electromagnética como son los fenómenos de difracción, refracción, absorción, etc. El modelo requerido debe ser adaptable de manera sencilla a un entorno urbano o rural, independiente de la geografía y que permita sacar conclusiones válidas sobre la predicción de señal sobre la zona a cubrir.

Con el surgimiento de las redes inalámbricas el problema de planificación de la red de radio ha tenido numerosas formas de resolución. En [7] se realiza una revisión general de diferentes metaheurísticas resolviendo el problema de RND y utilizan la función objetivo propuesta por Calegari et al. en [2, 3]. El estudio pretende ofrecer una base de referencia confiable sobre un amplio espectro de algoritmos y medidas precisas de comparación de la eficiencia, confiabilidad y rapidez de las diferentes técnicas aplicadas a la resolución del RND. Se dan referencias de uso de AGs con implementaciones híbridas en [18], paralelas o multiobjetivo en [5, 6, 8, y 22]. Un ejemplo de esto se puede ver en [1] donde se modela el problema en su máxima complejidad práctica, basado una optimización combinatoria con restricciones multiobjetivo con un enfoque Pareto para encontrar el conjunto de redes no dominantes y diversificadas. En [20] se plantea el uso de una heurística para resolver el problema de RND, formada por tres fases

secuenciales. La primera fase ejecuta el preprocesamiento de las restricciones del problema, eliminando de esta forma un gran número de “malas” combinaciones. La segunda ejecuta una optimización utilizando Búsqueda Tabú (*Taboo Search*, TS) sobre un espacio de búsqueda reducido. La tercera es una post optimización ejecutando la sintonía fina de los parámetros de las antenas. Uno de los primeros enfoques de un algoritmo genético (AG) no estándar es el denominado CHC. El CHC es un AG no tradicional que utiliza una estrategia de selección muy conservadora: elige siempre a los mejores individuos para formar parte de la nueva población. Aplica además un operador de recombinación (HUX, del inglés Half Uniform Crossover), que produce una descendencia lo más diferente posible a ambos padres; el CHC incorpora un mecanismo para reiniciar el algoritmo ante una condición de convergencia prematura. También introduce un sesgo para evitar el cruce de individuos similares.

El algoritmo CHC también se aplicó al problema de diseño de la red de radio y obtuvieron buenos resultados en [21] y además se destaca el poco uso de este algoritmo. En [17] se presenta una versión de CHC multiobjetivo que en lugar de ordenar las soluciones en base a un valor escalar (el *fitness*), el CHC multiobjetivo (Mo-CHC) las soluciones se ordenan mediante un método de ranking y *crowding*.

Líneas de investigación y desarrollo

La popularización de los servicios inalámbricos de comunicación (telefonía, internet, etc.) es cada vez mayor en la actualidad. La demanda hace que los usuarios exijan la disponibilidad de estos servicios en todo momento y en cualquier lugar. El costo de los equipamientos para brindar

el servicio con la calidad adecuada es elevado por lo tanto es fundamental realizar la correcta selección de los emplazamientos geográficos que permitan lograr una cobertura óptima de señal minimizando el uso de equipamientos.

Esta tarea se la denomina diseño de la red de radio (RND) y es un problema NP-duro de optimización, por lo tanto, es factible de ser tratado con metaheurísticas. Dados los buenos resultados obtenidos por el algoritmo CHC aplicado a este problema, en el presente trabajo se continuó la investigación generando seis versiones del algoritmo denominadas: QCHC-RE, QCHC-TE, QCHC-ML, QCHCMD, QCHC-MCB y QCHC-ILS. Las tres primeras versiones se basaron en el uso de diferentes métodos de selección poblacional y el resto de las versiones en variaciones del método de sacudida.

Además, propone aportes relacionados con comunicaciones, metaheurísticas, modelado de terreno para ser representado computacionalmente y se muestran los resultados obtenidos por cada uno de los algoritmos junto con el análisis estadístico de los mismos.

Resultados obtenidos/esperados

En cuanto a los resultados de ésta línea de investigación y desarrollo se obtuvieron hasta el momento los siguientes objetivos:

- Construcción de instancias de prueba artificiales para el problema de RND.
- Optimización de la función de evaluación basada en la eficiencia de cobertura
- Desarrollo de nuevas variantes del algoritmo CHC con respecto a los

mecanismos de selección y de sacudida.

- Construcción y prueba de instancias reales en los yacimientos de zona norte de Santa Cruz incluyendo restricciones topográficas del terreno.

Estos avances del proyecto son detallados en Molina et al [9, 10, 11, 12, 13, 14, y 15, y 16].

En trabajos futuros también se podrá abordar escenarios con un modelo de cobertura más específico, diferentes formas del lóbulo de radiación paralelización de los algoritmos secuenciales desarrollados.

Formación de recursos humanos

Un integrante del proyecto está desarrollando su tesis de Doctorado orientada a esta línea de investigación.

Un integrante del proyecto presentó y defendió su tesis de Maestría.

Referencias

[1] E. Alba and F. Chicano, On the behavior of parallel genetic algorithms for optimal placement of antennae in telecommunications, *Int. J. Found. Comput. Sci.*, vol. 16, pp. 86–90, 2005.

[2] P. Calegari, F. Guidec, P. Kuonen, and D. Wagner. Genetic Approach to Radio Network Optimizations for Mobile Systems. In *Proceedings 47th IEEE Conference on Vehicular Technology*, volume 2, pages 755-759, 1997. (a)

[3] P. Calegari, F. Guidec, P. Kuonen, and D. Kobler, *Parallel Island Based Genetic Algorithm for Radio Network Design*, J.

[4] *Parallel Distrib.Comput.*, vol. 47, no. 1, pp. 86–90, 1997.(b)

[5] G. Celli, E. Costamagna, and A. Fanni, *Genetic Algorithms for Telecommunication Network Optimization*, presented at *IEEE Int. Conf. Syst., Man and Cybernetics*, 1995.

[6] C. M. Fonseca and P. J. Fleming, *Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization*, in *Proc. 5th Int. Conf. Genetic Algorithms*, 1993, pp. 416–423.

[7] S. Mendes, G. Molina, M. Vega-Rodríguez, J. Gómez-Pulido, Y. Sáez, G. Miranda, C. Segura, E. Alba, P. Isasi, C. León, and J. Sánchez-Pérez, *Benchmarking a Wide Spectrum of Metaheuristic Techniques for the Radio Network Design Problem*. *EEE Transactions on Evolutionary Computation*, vol. 13, no. 5, October 2009.

[8] H. Meunier, E. G. Talbi, and P. Reininger, *A Multiobjective Genetic Algorithm for Radio Network Optimization*, presented at *Congr. Evol. Comput.*, 2000.

[9] D. Molina, D. Pandolfi, y A. Villagra. Aplicación y evaluación de diferentes algoritmos genéticos canónicos en el diseño eficiente de redes de radio frecuencia en comunicaciones inalámbricas. *Revista de Informes Científicos y Técnicos de la Universidad Nacional de la Patagonia Austral*. Vol.5, núm. 3 (2013). pag 135-161.

[10] D. Molina, D. Pandolfi, y A. Villagra. Diseño de coberturas de señales de radio frecuencia aplicando metaheurísticas. *III Encuentro de Investigadores de la Patagonia Austral*. ISBN 978-987-3714-00-9. 3era Edición.

- [11] D. Molina, D. Pandolfi, A. Villagra, y G. Leguizamón. Diseño eficiente de redes de radio frecuencia con algoritmos CHC en comunicaciones inalámbricas. Actas del 2 Congreso Nacional de Ingeniería Informática/Sistemas de Información, CoNaIISI (2014), pag. 125-133. ISSN: 2346-9927
- [12] D. Molina, D. Pandolfi, A. Villagra, y G. Leguizamón. Metaheurísticas aplicadas a la optimización de cobertura de señales de radio frecuencia con un modelo de propagación adaptable. XVI Workshop de Investigadores en Ciencias de la Computación (WICC 2014), p. 109-114, Ushuaia, Argentina.
- [13] D. Molina, D. Pandolfi, A. Villagra, y G. Leguizamón. Applying CHC Algorithms on Radio Network Design for Wireless Communication. Computer Science & Technology Series, XX Argentine Congress of Computer Science, Selected papers (2015), pag 27-37. ISBN 978-987-1985-71-5.
- [14] D. Molina, D. Pandolfi, A. Villagra, y G. Leguizamón. Variantes del algoritmo CHC para proyectar redes de radio frecuencia en comunicaciones inalámbricas. Revista de Informes Científicos y Técnicos de la Universidad Nacional de la Patagonia Austral. Vol.7, núm. 2 (2015). pag 228-248.
- [15] D. Molina, D. Pandolfi, A. Villagra, y G. Leguizamón. Hibridación del Algoritmo CHC aplicado al Cálculo de Coberturas en Redes Inalámbricas en Yacimientos Petroleros. 12 Congreso Interamericano de Computación Aplicado a la Industria de Procesos CAIP 2015. ISBN 978-958-8791-81-4.
- [16] D. Molina, D. Pandolfi, A. Villagra, y G. Leguizamón. Metaheurísticas aplicadas a la optimización de cobertura de señales de radio frecuencia para telemetrías en yacimientos petroleros. XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015), Salta, Argentina.
- [17] A. J. Nebro, G. Alba, Enrique y Molina, F. Chicano, and J. J. Luna, Francisco y Durillo. Optimal antenna placement using a new multi-objective chc algorithm. In Proceedings of the 9th annual conference on Genetic and evolutionary computation, pages 876–883. ACM, 2007.
- [18] A.R. Penin. Sistemas SCADA. Marcombo, (2012).
- [19] E. G. Talbi, S. Cahon, and N. Melab, “Designing cellular networks using a parallel hybrid metaheuristic on the computational grid,” *Comput. Commun.*, vol. 30, no. 4, pp. 698–713, 2007.
- [20] K. Tutschku, N. Gerlich, and P. Tranga An integrated Approach to Cellular Network Planning, Institute of Computer Science, University of Wurzburg, 1995.
- [21] M. Vega-Rodríguez, J. Gómez-Pulido, E. Alba, D. Vega-Pérez, S. Priem-Mendes, G. Molina, Evaluation of Different Metaheuristics Solving the RND Problem, *EvoWorkshops 2007, LNCS 4448*, pp. 101–110, 2007.
- [21] S. Watanabe, T. Hiroyasu, and M. Mikiand, Parallel Evolutionary Multicriterion Optimization for Mobile Telecommunication Networks Optimization, presented at Eurogen 2001—Evol. Methods Design, Optimisation Control with Applicat. Ind. Problems Conf., Athens, Greece, 2001.

Desarrollo de herramientas para la operabilidad de procesos productivos

Enrique E. Tarifa; Sergio L. Martínez; Samuel Franco Domínguez; Susana A. Chalabe; Álvaro F. Núñez

Facultad de Ingeniería / Universidad Nacional de Jujuy / CONICET
Ítalo Palanca N°10 / S. S. de Jujuy / Provincia de Jujuy / Tel. 0388-4221591
eetarifa@fi.unju.edu.ar; smartinez@fi.unju.edu.ar; sfrancodominguez@gmail.com;
susana.chalabe@gmail.com; alfanunez@yahoo.com.ar

Resumen

El presente proyecto tiene como objetivo general desarrollar herramientas para la operabilidad de procesos productivos. Para ello se desarrollarán, adaptarán y aplicarán herramientas propias de la Ingeniería de Procesos. Los objetivos particulares del proyecto propuesto son los siguientes: 1) Desarrollar métodos para determinar y aumentar la flexibilidad de los procesos productivos; 2) Desarrollar sistemas de control avanzados para aumentar la controlabilidad y robustez de los sistemas productivos —en especial, se emplearán técnicas de Inteligencia Artificial—; 3) Desarrollar sistemas de diagnóstico de fallas para aumentar la operabilidad de los procesos productivos; 4) Desarrollar material para la enseñanza de Ingeniería Química (simuladores, optimizadores, sistemas de control, publicaciones y apuntes). Los procesos a estudiar serán procesos productivos vinculados a la industria química, a la industria de alimentos y a todo proceso productivo de la región que pueda beneficiarse con las herramientas a desarrollar.

Palabras clave: Operabilidad, Flexibilidad, Controlabilidad, Confiabilidad, Robustez.

Contexto

El proyecto en cuestión se desarrolla en el ámbito de la Facultad de Ingeniería de la Universidad Nacional de Jujuy (UNJu), con la dirección del Dr. Enrique Tarifa (investigador de CONICET) y la codirección del Mg. Ing. Sergio Martínez.

Las actividades de investigación son desarrolladas por el grupo de investigación *IngProAr* en la sede del Instituto de Tecnología Minera e Industrial (InTeMI) de la Facultad de Ingeniería de la UNJu. Este grupo de investigación tiene carácter multidisciplinario, ya que cuenta con miembros provenientes de las siguientes cátedras: Simulación y Optimización, Ingeniería de Procesos, Introducción a la Informática, e Inteligencia Artificial.

El financiamiento del proyecto es provisto por la Secretaría de Ciencia y Técnica y Estudios Regionales (SeCTER) de la Universidad Nacional de Jujuy (UNJu). El proyecto pertenece a la categoría A, y está inserto en el Programa de Incentivos dependiente de la Secretaría de Políticas Universitarias (SPU) del Ministerio de Educación de la Nación con el código 08/D138, desde 2014.

Por otra parte, este proyecto forma parte del Programa de Investigación “Desarrollo de Sistemas de Soporte a la Toma

de Decisiones” que también es dirigido por el Dr. Enrique Tarifa y codirigido por el Mg. Ing. Sergio Martínez. Los proyectos que pertenecen a este programa también tiene como lugar de trabajo la Facultad de Ingeniería de la UNJu.

Introducción

El presente proyecto es continuación del proyecto “Desarrollo de sistemas de apoyo para la toma de decisiones en procesos industriales”, que fue financiado por SeCTER (Secretaría de Ciencia y Técnica y Estudios Regionales) y la Facultad de Ingeniería de la UNJu, y reconocido por el Programa de Incentivos con el código 05/D106. Durante el desarrollo de dicho proyecto fue necesario avanzar sobre temas que son propios de la operabilidad de los procesos. Es así que se desarrollaron trabajos de diagnóstico de fallas [1, 2], control inteligente [3, 4], flexibilidad [5], simulación y optimización de procesos industriales [6, 7]. Los resultados logrados en estas áreas fueron publicados en congresos y revistas internacionales [1-7]. Durante esta investigación, se determinó que los citados temas tienen gran importancia tanto en el ámbito académico como en el industrial; por tal motivo, el presente proyecto está orientado a profundizar la investigación de esos temas. Es decir, el proyecto propuesto continúa con la línea de investigación hasta aquí realizada, y se enfoca en el área de operabilidad. Es también objetivo del presente proyecto, producir material para la enseñanza de la Ingeniería Química. Estos materiales serán libros, simuladores, publicaciones, elementos multimedia y todo lo que se considere de utilidad para esta tarea.

La operabilidad de un proceso productivo comprende las siguientes propieda-

des del mismo: flexibilidad, controlabilidad, confiabilidad y robustez.

Debido a que cada una de las propiedades enunciadas es importante por sí misma, existen trabajos que se centran en cada una de ellas. Los estudios de flexibilidad suelen enfocarse en la definición y posterior determinación de índices que capturen apropiadamente esta característica del proceso. Entre las acciones a tomar para aumentar la flexibilidad de un proceso está la modificación del diseño del mismo. En el campo de la controlabilidad, los estudios se orientan al planteo y verificación de condiciones que aseguren la controlabilidad de los procesos. Entre las acciones a tomar para mejorar la controlabilidad del proceso está la modificación del diseño y la implementación de un adecuado sistema de control. En el campo de la confiabilidad, los estudios tienen como objetivo determinar la probabilidad de ocurrencia de fallas y sus consecuencias. En este caso, las acciones a tomar son preventivas (destinadas a disminuir la probabilidad de falla y sus consecuencias) y paliativas (producida la falla, se toman las medidas necesarias para atenuar sus efectos). En el campo de la robustez, los trabajos se concentran en el diseño de sistemas de control apropiados que se comporten adecuadamente en un rango amplio de condiciones de operación. Finalmente, debido a que las acciones que se tomen para mejorar una propiedad también afectan a las restantes, existen trabajos que estudian simultáneamente dos o más de las citadas propiedades.

Es de destacar que los actuales estudios de operabilidad involucran cada vez más el uso de técnicas de Inteligencia Artificial (redes neuronales, sistemas expertos con lógica *fuzzy*, entre otras), siendo los sistemas de control inteligentes los mejores exponente de esta situación. También se han aplicado con éxito estas técnicas para mejorar el diseño de los

procesos, recurriendo en especial a estrategias evolutivas.

Líneas de Investigación, Desarrollo e Innovación

El proyecto propuesto se enmarca en las siguientes líneas prioritarias de la UNJu (Res. CS N°168/93):

- **Desarrollo Económico – Social Regional Sustentable:** Las herramientas a desarrollar en el proyecto servirán para aumentar la operabilidad de los procesos productivos de la región. En consecuencia, se favorecerá el desarrollo económico de la región.

También se enmarca en las siguientes líneas prioritarias de la Facultad de Ingeniería (Res. FI N°071/98):

- **Línea 1 - Educación:** El proyecto en consideración producirá material para la enseñanza de Ingeniería Química. Además del material a publicar, se tendrá también documentos electrónicos y multimedia que serán volcados en las aulas virtuales de las materias que tienen a cargo los integrantes del grupo *IngProAr*.
- **Línea 3 - Ingeniería de Procesos:** El proyecto en consideración se enmarca principalmente en esta línea prioritaria ya que la operabilidad de los procesos productivos es una importante área de la Ingeniería de Procesos.

Durante el año 2013, la Facultad de Ingeniería de la UNJu trabajó en la reformulación de las líneas prioritarias. De aprobarse el borrador actualmente existente, el proyecto propuesto se enmarca principalmente en la siguiente línea:

- **Línea 3 - El estudio de procesos específicos o integrados que con-**

tribuyan a la cadena de valor de los productos obtenidos: Esta línea comprende los siguientes temas: Análisis, diseño y síntesis óptima de procesos; Simulación y optimización de procesos; Operación y mantenimiento de procesos; Supervisión y control de procesos; Sistemas de apoyo a la toma de decisiones en procesos; Integración de masa y energía; Análisis de la Performance de procesos. Los resultados a obtener con el proyecto propuesto están vinculados a todos los temas citados.

Finalmente, la aplicación de las herramientas a desarrollar para aumentar la operabilidad de procesos productivos implicará aumentos en la producción y en la calidad de la misma, con los consecuentes efectos positivos sobre la sociedad.

Por todos los motivos expuestos, se considera relevante para el país y la provincia el objeto de investigación del proyecto propuesto.

Resultados y Objetivos

El presente proyecto tiene como objetivo general desarrollar herramientas para la operabilidad de procesos productivos. Para ello se desarrollarán, adaptarán y aplicarán herramientas propias de la Ingeniería de Procesos.

Los objetivos particulares del proyecto propuesto son los siguientes:

- 1) Desarrollar métodos para determinar y aumentar la flexibilidad de los procesos productivos.
- 2) Desarrollar sistemas de control avanzados para aumentar la controlabilidad y robustez de los sistemas productivos. En especial, se emplearán técnicas de Inteligencia Artificial.

- 3) Desarrollar sistemas de diagnóstico de fallas para aumentar la operabilidad de los procesos productivos.
- 4) Desarrollar material para la enseñanza de Ingeniería Química: simuladores, optimizadores, sistemas de control, publicaciones y apuntes.

Los procesos a estudiar serán procesos productivos vinculados a la industria química, a la industria de alimentos y a todo proceso productivo de la región que pueda beneficiarse con las herramientas a desarrollar.

En los dos años de ejecución del proyecto se obtuvieron resultados que dieron lugar a trabajos presentados en congresos, publicaciones en revistas científicas y capítulos de libros. Además, se desarrollaron programas informáticos para la simulación, optimización y control de los procesos estudiados.

Formación de Recursos Humanos

A continuación se detallan las actividades desarrolladas en formación de recursos humanos en el marco del proyecto en cuestión:

Becas de postgrado

- Ing. Álvaro F. Núñez, con beca doctoral de la ANPCyT, desde 2009. Cursa el Doctorado Regional en Ciencia y Tecnología de Alimentos en la Facultad de Ingeniería de la UNJu. Tesis doctoral “Simulación, optimización y control de procesos de la industria alimentaria de la región”. Director: Dr. Enrique Tarifa.
- Ing. Leonel Benítez, con beca doctoral de CONICET, desde 2011. Cursa el Doctorado en Ingeniería en la Facultad de Ingeniería de la UNSa. Tesis doctoral “Desarrollo de procedimientos óptimos de operación para el sector de ajuste

de punto del rocío en plantas de acondicionamiento de gas natural”. Director: Dr. Enrique Tarifa.

- Ing. Juan Pablo Gutiérrez, con beca doctoral de CONICET, desde 2012. Cursa el Doctorado en Ingeniería en la Facultad de Ingeniería de la UNSa. Tesis doctoral “Optimización del diseño, condiciones de operación y sistema de control del sector de endulzamiento de una planta de acondicionamiento de gas natural”. Director: Dr. Eleonora Erdmann, Codirector: Dr. Enrique Tarifa.
- Ing. Karina Alejandra Palma, desde 2015. Cursa la Maestría en Gestión Ambiental en la UCASAL. Tesis de maestría “Diseño de un Sistema de Gestión Ambiental para la empresa de tratamiento de líquidos cloacales El Cadillal”, Director: Dr. Enrique Tarifa.
- Ing. Lara Valeria Lescano Farías, desde 2013. Cursa el Doctorado en Ciencia y Tecnología de los Alimentos de la UNSE. Tesis doctoral “Secado en lecho de chorro bidimensional para la deshidratación de proteínas del plasma y porción globina de sangre bovina”. Director: Dra. Eve Liz Coronel, Codirector: Dr. Enrique Tarifa.

Becas en investigación

- Andrea Celeste Nievez, estudiante de Ingeniería Química de la UNJu, con beca de Estímulo a las Vocaciones Científicas, CIN, 2013-2015. Título del plan de trabajo “Optimización de una planta industrial de litio”. Director: Dr. Enrique Tarifa.
- Cristian David Yurquina, estudiante de Ingeniería Química de la

UNJu, con beca de Estímulo a las Vocaciones Científicas, CIN, 2015-2016. Título del plan de trabajo “Optimización del tren de piletas de concentración de una planta industrial de litio”. Director: Dr. Enrique Tarifa.

- Lautaro Acosta, estudiante de Ingeniería Química de la UNJu, con beca de Estímulo a las Vocaciones Científicas, CIN, 2015-2016. Título del plan de trabajo “Modelado y control inteligente de un proceso de concentración de Litio mediante pozas solares”. Director: Mg. Ing. Sergio Luis Martínez.

Referencias

1. Martínez S.L., Franco Domínguez S., Tarifa E.E., Sánchez Rivero V.D., Azar M.A., “Desarrollo de un Sistema de Diagnóstico de Fallas para Equipos de Procesos basado en Estados Anormales Caracterizados”, XIII Workshop de Investigadores en Ciencias de la Computación WICC 2011, 5 y 6 de mayo de 2011, RedUNCI, Universidad Nacional de Rosario, Rosario, Santa Fe, ISBN 978-950-673-892-1, 772-776, 2011.
2. Martínez S.L., Tarifa E.E., Núñez A.F., “Comparación de Desempeño de Controladores Fuzzy y Controladores Proporcionales”, Avances en Ciencias e Ingeniería, ISSN: 0718-8706, 3 (3), 91-101, 2012.
3. Martínez S.L., Tarifa E.E., Sánchez Rivero V.D., “Control neuronal tipo MIMO aplicado a un mezclador de corrientes líquidas”, VII Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA, 13 y 14 de octubre de 2011, San Fernando del Valle de Catamarca, Catamarca, 2011.
4. Martínez S.L., Tarifa E.E., Sánchez Rivero V.D., “Derivación de Reglas de Control MIMO para un Controlador Fuzzy sin Conocimiento Experto”, Cuarto Simposio Internacional de Investigación: “La investigación en la Universidad: Experiencias Innovadoras en Investigación Aplicada”, 19 y 21 de octubre de 2011, San Salvador de Jujuy, Jujuy, 2011.
5. Tarifa E.E., Núñez A.F., Franco Domínguez S., “Estudio de flexibilidad de un evaporador doble efecto para jugo de tomate”, VII Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA, 13 y 14 de octubre de 2011, San Fernando del Valle de Catamarca, Catamarca, 2011.
6. Guzmán A.G., Luck D.A., Horowitz G.I., Vera C.R., Tarifa E.E., “Optimización de la Gestión del Hidrógeno en Refinería Luján de Cuyo”, 3er. Congreso Latinoamericano y del Caribe de Refinación, del 30 de octubre al 2 de noviembre de 2012, Buenos Aires, 2012.
7. Benítez L.A., Erdmann E., Tarifa E.E., “Desarrollo de procedimientos óptimos de operación para el sector de ajuste de punto del rocío en plantas de acondicionamiento de gas natural”, I Jornadas de Jóvenes Investigadores, 22 y 23 de noviembre de 2012, Facultad de Humanidades y Ciencias Sociales, UNJu, 2012.

Benchmarks para Problemas de Scheduling de Máquinas Paralelas Idénticas con Algoritmos Inteligentes

Gatica Claudia Ruth, Esquivel Susana Cecilia

LIDIC

Departamento de Informática/ FCFMyN

Universidad Nacional de San Luis

Ejército de los Andes 950 – Local 106, Tel: (266) 4420823

crgatica,esquivel@unsl.edu.ar

Resumen

En nuestro trabajo presentamos un conjunto de *benchmarks* para el problema de *scheduling* de paralelas idénticas sin restricciones. Hemos estudiado tal problema a través de cuatro funciones objetivo: *Maximum Tardiness (Tmax)*, *Average Tardiness (Tavg)*, *Total Weighted Tardiness (Twt)* y *Weighted Number of Tardy Jobs (Nwt)*. El conjunto de *benchmarks* se dividen en ocho escenarios de 125 instancias cada uno. Tales instancias del problema se construyeron en base a datos seleccionados de la *OR-Library* [6] correspondientes a problemas de tardanza ponderada. Se obtuvieron los valores de los óptimos conocidos o *benchmarks* mediante la aplicación de reglas de despacho y heurísticas conocidas en la literatura [7, 8] y luego se utilizaron dos algoritmos propuestos basados en búsqueda local: uno de ellos es una variación del algoritmo *Simulated Annealing (SA-explorador)*, el segundo algoritmo es de Vecindarios Variables (*VNS*) y el tercer algoritmo basado en

búsqueda poblacional conocido como *Discrete Differential Evolution (DDE)*.

Palabras clave: *benchmarks*, *scheduling*, máquinas paralelas idénticas, algoritmo *Simulated Annealing (SA)* y búsqueda en Vecindarios Variables (*VNS*), *Discrete Differential Evolution (DDE)*.

Contexto

La línea de investigación se enmarca en el paradigma de “Algoritmos Inteligentes”, en el proyecto de investigación: “Diseño de distribución de turbinas en parques eólicos usando métodos de optimización aproximados”. Tal proyecto esta en ejecución desde el 1 de enero de 2014 e inserto dentro del Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC).

Introducción

El problema en estudio es la planificación (*scheduling*) de máquinas idénticas paralelas con respecto a cuatro funciones objetivos: la máxima tardanza (*Maximum Tardiness: T_{max}*), la tardanza media

(Average Tardiness: T_{avg}), la tardanza pondera total (Total Weighted Tardiness: T_{wt}) y el número ponderado de tareas tardías (Weighted Number of Tardy Jobs: N_{wt}). La notación usada en la literatura [7] para describir el problema es una tripla: $(\alpha | \beta | \delta)$. El primer campo α describe el ambiente de máquinas, el segundo campo β indica las restricciones entre las tareas que son asignadas a las máquinas. Por último el tercer campo δ provee la o las funciones objetivo a ser optimizadas. Acorde a esta notación, nuestro problema se describe mediante $(P_m || \delta)$, el campo δ representa las funciones objetivo. Tal problema se ha considerado en [8] de complejidad NP-duro cuando $2 \leq m \leq n$ (m es el número de máquinas y n el número de tareas).

Líneas de Investigación, Desarrollo e Innovación

En los trabajos relacionados es común encontrar instancias de prueba para la planificación de máquina única, sin embargo para problemas de máquinas paralelas es casi imposible. Así que construimos nuestro propio conjunto de instancias de prueba a partir de los datos seleccionados correspondientes a un problema de tardanza ponderada tomada de la OR-Library [6]. Uno de los objetivos principales de nuestro trabajo es el estudio de algoritmos basados en búsqueda local y poblacional, otro de los objetivos principales es la construcción de *benchmarks* para instancias de prueba del problema de máquinas idénticas paralelas a través de la aplicación de los algoritmos estudiados.

Resultados y Objetivos

Se realizó un estudio de parámetros de los algoritmos propuestos SA-explorador,

VNS y DDE mediante el método de cuadrados latinos (LH) de la literatura [2], se eligieron 20 y 30 configuraciones diferentes de parámetros y se realizó un estudio estadístico mediante el software CONTROLTEST que provee varios *test* estadísticos de los cuales elegimos el *test* de Friedman [2] para clasificar los resultados, esto es, elegir la mejor configuración c_i y luego realizar comparaciones de a pares y analizar diferencias significativas.

Los resultados alcanzados fueron estudiados con varias métricas de evaluación propuestas y han sido plasmados en reportes técnicos. Los escenarios del conjunto de instancias son los siguientes: **I:** $n = 100$ y $m = 5$; **II:** $n = 100$ y $m = 2$; **III:** $n = 40$ y $m = 5$; y **IV:** $n = 40$ y $m = 2$, **V:** $n=100$ y $m=15$, **VI:** $n=100$ y $m=30$, **VII:** $n=40$ y $m=15$; **VIII:** $n=40$ y $m=30$, donde n es el número de tareas y m es el número de máquinas. Nuestro objetivo es presentar un conjunto de *benchmarks* que estén disponibles vía web para la investigación de problemas de planificación de máquinas paralelas idénticas en el futuro.

Formación de Recursos Humanos

El presente trabajo corresponde al plan de trabajo de la tesis doctoral: “Desarrollo y Aplicación de Metaheurísticas basadas en Inteligencia Computacional para resolver problemas de Planificación de Máquinas Paralelas Idénticas”, de la Carrera de Posgrado: Doctorado en Ciencias de la Computación. El trabajo se divide en dos etapas. La Etapa I que involucra el estudio de algoritmos inteligentes basados en búsqueda poblacional y local. La Etapa II comprende la aplicación de los algoritmos estudiados para construir con conjunto de *benchmarks* para instancias del problema y colocarlos a disposición de interesados en la investigación del

problema de *scheduling* de máquinas idénticas paralelas vía web.

Referencias

[1] E.G. Talbi, "Metaheuristics from design to implementation", by John Wiley & Sons, Canada, 2009.

[2] Joaquín Derrac and Salvador García and Daniel Molina and Francisco Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms", Swarm and Evolutionary Computation, 2011.

[3] T. Bartz-Beielstein, "Experimental Research in Evolutionary Computation", The New Experimentalism, Springer, 2006.

[4] Liao C. J., Cheng C. C., "A variable neighborhood search for minimizing single machine weighted earliness and tardiness with common due date", Computers and Industrial Engineering, 52(4), 404-413, 2007.

[5] Hiba Yahyaoui, Saoussen Krichen, Bilel Derbel, El-Ghazali Talbi, "A Variable Neighborhood Descent for solving the Single Machine Total Weighted Tardiness Problem", IEEE, 2013.

[6] J.E. Beasley, ORLIB, <http://people.brunel.ac.uk/mastjbjb/jeb/orlib/wtinfo.html>.

[7] Pinedo M., "Scheduling: Theory, Algorithms and System", Prentice Hall, 1995.

[8] Morton T. and Pentico D., "Heuristic Scheduling Systems, John Wiley and Sons", 1993, New York

[9] Hong-yan Sang., "A Discrete Differential Evolution Algorithm for Lotstreaming Flow Shop Scheduling Problems", 2010 Sixth International Conference on Natural Computation (ICNC 2010).

Toma de decisiones y aprendizaje en agentes artificiales inteligentes

Edgardo Ferretti, Cecilia Sosa Toranzo, Guillermo Aguirre,
Juan Martín Loyola, Leticia Cagnina*, Marcelo Errecalde

Laboratorio de Investigación y Desarrollo en Inteligencia Computacional
Departamento de Informática, Universidad Nacional de San Luis
e-mails de contacto: {ferretti,lcagnina,merreca}@unsl.edu.ar

Resumen

Este artículo describe, brevemente, las tareas de investigación y desarrollo que se están llevando a cabo en la línea de investigación “Toma de decisiones y aprendizaje automático” en el marco del proyecto “Aprendizaje automático y toma de decisiones en sistemas inteligentes para la Web”. Esta línea se centra en la formalización, diseño y desarrollo de modelos formales y mecanismos para la implementación de sistemas basados en Agentes Inteligentes. En particular, se hace especial hincapié en aquellos procesos y mecanismos que proveen las capacidades fundamentales usualmente asociadas al comportamiento “inteligente”, tales como el aprendizaje automático, el razonamiento y la toma automática de decisiones, la planificación de acciones para alcanzar un objetivo (planning) y la comunicación entre agentes, entre otros.

Las temáticas incluyen el uso de argumentación en la toma de decisiones, arquitecturas de razonamiento práctico, agentes BDI, agentes basados en Procesos de Decisión de Markov, aprendizaje, votación y argumentación en arquitecturas de agentes inteligentes, comunicación entre agentes deliberativos, toma de decisiones “anticipada”, etc.

Palabras clave: Aprendizaje Automático, Toma de Decisiones Automática, Procesos de Decisión Secuencial, MDPs,

*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Arquitecturas BDI, Argumentación, Votación, Comunicación

Contexto

La línea de investigación “Toma de decisiones y aprendizaje automático” es una de las tres líneas del proyecto “Aprendizaje automático y toma de decisiones en sistemas inteligentes para la Web”, un nuevo proyecto que será presentado este año como continuación del Proyecto de Investigación Consolidado (PROICO) titulado “Herramientas y mecanismos para la toma de decisiones en agentes inteligentes artificiales”. Este último proyecto, aprobado por evaluadores externos a la UNSL, se desarrolla en el *Laboratorio de Investigación y Desarrollo en Inteligencia Computacional* (LIDIC) de la UNSL y ha sido financiado en forma directa por la UNSL (proyecto PROICO 30312) y en forma indirecta por: (a) el Programa de Incentivos (Nro. 22/F237), (b) la Comisión Europea de Investigación e Innovación, a través del programa Marie Curie Actions: FP7 People 2010 IRSES, (c) el CONICET, a través de un investigador asistente y becas: dos de Doctorado y dos de Post-Doctorado asignadas a integrantes del proyecto y (d) el Consejo Nacional de Ciencia y Tecnología (CONACYT-México) y otros organismos científicos del gobierno mexicano, en los que distintos integrantes han participado en tres proyectos de investigación como colaboradores externos.

Este proyecto posee además otras dos líneas de investigación denominadas: “Aplicaciones” y “Minería de textos y de la Web”; la primera enfocada en el uso del aprendizaje automático en psicología, educación y el cuidado de la salud y la segunda vinculada a la ingeniería del lenguaje natural, incluyendo temas como el Procesamiento del Lenguaje Natural (PLN) y la Lingüística Computacional, la Minería de Textos y de la Web y la recuperación de información de la Web. Es claro en este contexto, que muchos problemas y aplicaciones intersectan los alcances de más de una de las líneas de este proyecto, lo cual involucra un trabajo integrado y coordinado permanente a los fines de optimizar los recursos disponibles para la obtención de los objetivos propuestos.

Introducción

La *toma de decisiones* es un campo de investigación muy activo en áreas muy diversas, como la Filosofía, Economía, Psicología, etc. y donde ha sido estudiada usando las metodologías científicas propias de cada disciplina. En particular, en Ciencias de la Computación, los problemas de toma de decisión han sido abordados mayormente desde el campo de investigación relativo a la Inteligencia Artificial, y donde el concepto de *agente artificial inteligente* (AAI) juega un rol protagónico. No obstante, no existe actualmente un consenso en la definición de AAI, y tal vez la acepción más comúnmente aceptada es la dada en [26]; donde un agente es un sistema computacional (basado en hardware o software) que es *autónomo, social, reactivo y pro-activo*. Otra característica a considerar en ambientes no determinísticos, desconocidos y altamente dinámicos como lo son la mayoría de los problemas de la vida real, es la capacidad de **aprendizaje y adaptación** que tienen los agentes artificiales “embebidos” en este tipo de ambientes. Como es bien sabido, este tipo de capacidades son cruciales para enfrentar la potencial falta de conocimiento inicial de los agentes como también para actualizar el conocimiento que se tiene sobre un ambiente y/o agentes que cambian a lo largo del tiempo.

Proveer a un agente de las capacidades necesarias para exhibir el comportamiento flexible descrito anteriormente, no es una tarea trivial. De hecho, una de las áreas de investigación más activa en el ámbito de agentes ha sido la definición de modelos y arquitecturas de agentes que intentan dar una respuesta a este problema. Existen arquitecturas que se han concentrado en el aspecto de la reactividad [1] y otras en cambio que han privilegiado los mecanismos de deliberación y planificación necesarios para proveer un comportamiento pro-activo [14]. Sin embargo, existe actualmente consenso de que cualquier arquitectura realista de agente, debería proveer un soporte adecuado para todas estas capacidades. Las arquitecturas híbridas [10] y las basadas en comportamientos [15] han intentado lograr un balance adecuado entre pro-actividad y reactividad. Sin embargo, las arquitecturas que mayor atención han recibido para este propósito son las denominadas **BDI** [6]. Similarmente, aquellas personas que se han orientado a los enfoques basados en utilidad, han tomado a los **Procesos de Decisión Markov** [4] y los juegos estocásticos [5], como los modelos principales para diseñar e implementar AAI.

En todas las arquitecturas mencionadas previamente, el fin último de los procesos internos y de razonamiento de los agentes, es seleccionar las acciones que éste realizará. En este contexto, estos procesos orientados hacia las acciones suelen ser referenciados como de *razonamiento práctico* [25], para diferenciarlo del razonamiento teórico, que sólo afecta las creencias del agente sobre el mundo. Sin embargo, desde un punto de vista más general, estas propuestas también pueden ser consideradas como modelos y arquitecturas para la *toma de decisiones* de AAI.

Por otra parte, en un *sistema multi-agente* (SMA) [25] un conjunto de agentes interactúa para conseguir algún objetivo difícil de lograr de forma individual. Usualmente, en este tipo de sistemas, cada agente posee información incompleta sobre su entorno y sobre sí mismo, teniendo además capacidades limitadas. El entorno es dinámico y cambiante, viéndose afectado por las acciones de todos los a-

gentes. Es por eso que la comunicación entre agentes emerge como un aspecto sumamente importante, ya que la misma es una actividad netamente social que los agentes usan como parte de un plan para conseguir sus metas. Cada acto comunicativo [2] tiene su razón de ser, y durante su accionar, los agentes deben tomar una decisión acerca del acto del habla más adecuado para cada situación, lo que es una tarea compleja.

En este contexto, ha surgido un nuevo paradigma para la próxima generación de sistemas distribuidos y abiertos, donde las interacciones entre agentes computacionales están basadas en el concepto de *acuerdo*. En el pasado, dicho concepto, fue dominio de estudio para filósofos y sociólogos, y fue aplicado solamente a la sociedad humana. Sin embargo, hoy en día, el acuerdo y todos los procesos y mecanismos involucrados en lograr acuerdos entre diferentes agentes son también tema de análisis desde perspectivas orientadas a la tecnología, dando lugar a sistemas computacionales donde las componentes interactúan usando protocolos de argumentación [3, 11], votación [8, 21], negociación [16, 12] y mercado [17], entre otros. Dicho enfoque constituye la visión denominada **Tecnologías de Acuerdo** (en inglés, *agreement technologies* (AT)) [13] que engloba métodos, plataformas y herramientas para definir, especificar y verificar tales sistemas. Este nuevo paradigma ha quedado de manifiesto con la creación de consorcios de investigación europeos especializados en la temática¹ como así también en la organización de Conferencias Internacionales específicas de Tecnologías de Acuerdo.

Si bien las AT han recibido una atención considerable en el área de SMA, su aplicación dentro de las arquitecturas y modelos de agentes individuales para la toma de decisiones racional es un área que no ha recibido, hasta ahora, la necesaria atención. Más allá de trabajos aislados sobre el uso de votación en sistemas basados en comportamiento [15] y algunos esfuerzos por incorporar argumentación y/o votación en arquitecturas BDI [18, 21] existe una falencia notoria de propuestas tendientes a sistematizar el uso de estas

tecnologías en modelos de toma de decisión ya establecidos, como las arquitecturas BDI y las arquitecturas basadas en utilidades. Este no es un dato secundario si consideramos que el diseño modular de los agentes, o bien la existencia de múltiples criterios o preferencias personales que pueden entrar en conflicto, conducen naturalmente a situaciones que podrán ser modeladas como un SMA que funciona “dentro de la cabeza del agente”.

Por otra parte, si bien el aprendizaje automático (en inglés, *Machine Learning* (ML)) ha sido ampliamente desarrollado en modelos como los MDPs, no existe una clara especificación de cómo estos procesos podrían llevarse a cabo en enfoques basados en arquitecturas BDI, o de qué manera ML y AT se podrían complementar y/o asistir mutuamente. De igual manera, y a excepción de algunos trabajos recientes en el área [19, 20], tampoco existen suficientes investigaciones dedicadas a analizar la relación entre los MDPs y las arquitecturas BDI, lo que permitiría en primer lugar, integrar las ventajas y atenuar o solucionar los problemas que cada una de ellas exhibe por separado.

Líneas de Investigación y desarrollo

En base a lo anterior, se puede apreciar la disposición de enfoques y modelos muy potentes para el desarrollo de AAI como los MDPs, las arquitecturas BDI, las AT y ML. Por otra parte, si bien la integración de algunos de estos enfoques ha comenzado a realizarse recientemente, no existe, de acuerdo a nuestro conocimiento, propuestas que avancen en una integración de todos ellos en forma simultánea y consideramos, que esto permitiría modelos efectivos, eficientes y flexibles para la toma de decisiones de alto nivel de AAI. Por lo tanto, es en esta problemática donde se pretende generar las principales contribuciones, teniendo como ejes principales de investigación: (1) Arquitecturas de Razonamiento Práctico, en particular los MDPs y las arquitecturas BDI; (2) Tecnologías de Acuerdo; y (3) Aprendizaje automático.

¹<http://www.agreement-technologies.org>

Resultados y objetivos

Las principales motivaciones surgen de la intención de dar respuestas a las falencias descritas previamente, y plantear contribuciones en el área a partir del cumplimiento de los siguientes objetivos: (1) Lograr un entendimiento más profundo de los distintos enfoques basados en AT y aprendizaje automático para su uso en distintas componentes y procesos de modelos de razonamiento práctico, como las arquitecturas BDI y algoritmos basados en MDP y sus variantes. Este aprendizaje de los enfoques, debería incluir la determinación y comparación de las propiedades formales y los aspectos computacionales que involucran estas técnicas y modelos. (2) Obtener, mediante un proceso adecuado de generalización, una arquitectura abstracta para la incorporación de AT y aprendizaje en modelos basados en MDPs y BDI. La idea es lograr un modelo y una arquitectura genérica donde estas componentes puedan ser fácilmente integradas e intercambiadas. En este sentido, nuestra idea es lograr un marco de trabajo genérico para la integración de AT dentro de las arquitecturas de razonamiento práctico al igual que el soporte para el aprendizaje.

En este contexto, ya se han logrado algunos avances. En [24] se presentó un marco que integra AT desde un punto de vista multicriterio para soportar la toma de decisiones internas de un agente BDI; a saber: selección de deseos conflictivos, selección de planes y reconsideración de intenciones. Además, para abordar tales aspectos se implementó un enfoque concreto basado en votación. Así este trabajo extiende una propuesta anterior basada en argumentación [22], y se demuestra su adecuación como marco abstracto al presentar una instanciación, con un mecanismo basado en votación, en más etapas de decisión internas, que las consideradas originalmente en [22]. Asimismo, en [23], se presenta una propuesta novedosa, donde la reconsideración de intenciones es planteada en términos de un modelo de elección dicotómico, lo que permite cambiar de forma flexible las intenciones con las que se ha comprometido el

agente dependiendo de cómo evoluciona su entorno.

El trabajo a futuro, incluye el uso de modelos inicialmente pensados para procesos de decisión secuencial (MDPs), en problemas de categorización de textos donde es crítico realizar la clasificación *tan pronto* como sea posible. Estos escenarios conocidos como “early text classification”, ya han sido abordados como MDPs en [7] y serían de suma importancia en problemas de detección de pedófilos y de bullying. Integrantes de nuestro grupo, ya han participado en estudios preliminares en esa dirección [9], y se dirige un trabajo final de Licenciatura en esta temática. Respecto al área de comunicación entre agentes, está previsto continuar con el estudio y experimentación de distintos modelos de comunicación considerando las ventajas y desventajas de los dos principales estrategias: el enfoque *mentalista*, que sigue la teoría de los actos del habla en base a las creencias del agente y el enfoque *social*, donde la coordinación de acciones se consigue a través de acuerdos sustentados en la confianza y los compromisos asumidos.

Formación de Recursos Humanos

Trabajos de tesis vinculados con las temáticas descritas previamente:

- Dos tesis Doctorales en ejecución (una con beca de la UNSL).
- Un trabajo final de Licenciatura.

Referencias

- [1] R. C. Arkin. *Behaviour-Based Robotics*. The MIT Press, 1998.
- [2] J. L. Austin. *How to Do Things with Words*. Harvard University Press, Cambridge, MA, Cambridge, MA, 1962.
- [3] J. Bentahar, R. Alam, and Z. Maaamar. An argumentation-based protocol for conflict resolution. In *Workshop on Knowledge Representation for Agents and Multi-Agent Systems*, 2008.

- [4] C. Boutilier, T. Dean, and S. Hanks. Decision-theoretic planning: Structural assumptions and computational leverage. *J. Artif. Intell. Res. (JAIR)*, 1999.
- [5] M. Bowling and M. Veloso. An analysis of stochastic game theory for multiagent reinforcement learning. Technical Report 00-165, Computer Science Department, Carnegie Mellon University, 2000.
- [6] M. Bratman, D. Israel, and M. Pollack. Plans and resource bounded reasoning. *Computational Intelligence*, 1988.
- [7] G. Dulac-Arnold, L. Denoyer, and P. Gallinari. Text classification: A sequential reading approach. *LNCS*, 661:411–423, 2011.
- [8] M. Errecalde, G. Aguirre, and F. González. Agentes y mecanismos de votación. In *X CACIC*, 2004.
- [9] H. J. Escalante, M. Montes-y-Gómez, L. V. Pineda, and M. L. Errecalde. Early text classification: a naive solution. *CoRR*, abs/1509.06053, 2015.
- [10] E. Gat. On three-layer architectures. In *Artificial Intelligence and Mobile Robots*, 1998.
- [11] A. Kakas, N. Maudet, and P. Moraitis. Layered strategies and protocols for argumentation-based agent interaction. In *Argumentation in Multi-Agent Systems*, pages 64–77, 2005.
- [12] H. J. Müller. Negotiation principles. In *Foundations of distributed artificial intelligence*. John Wiley & Sons, 1996.
- [13] S. Ossowski. *Agreement Technologies*. Law, governance and technology series. Springer, 2013.
- [14] M. E. Pollack. Planning technology for intelligent cognitive orthotics. In *6th International Conference on AI Planning and Scheduling*, 2002.
- [15] J. Rosenblatt. Damn: A distributed architecture for mobile navigation - thesis summary. In *JETAI*, 1995.
- [16] J. S. Rosenschein and G. Zlotkin. *Rules of Encounter - Designing Conventions for Automated Negotiation among Computers*. The MIT Press, 1998.
- [17] T. Sandholm. Distributed rational decision making. In *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. The MIT Press, 1999.
- [18] F. Schlesinger, E. Ferretti, M. Errecalde, and G. Aguirre. An Argumentation-based BDI Personal Assistant. In *23rd IEA-AIE*. Springer, 2010.
- [19] G. I. Simari. *Markov Decision Processes and the Belief-Desire-Intention Model: Bridging the Gap for Autonomous Agents*. Springerbriefs in Computer Science. Springer New York, 2011.
- [20] G. I. Simari and S. Parsons. On the relationship between MDPs and the BDI architecture. In *AAMAS*, 2006.
- [21] C. Sosa, F. Schlesinger, E. Ferretti, and M. Errecalde. Integrating a voting protocol within an argumentation-based BDI system. In *XVI CACIC*, 2010.
- [22] C. Sosa-Toranzo, M. Errecalde, and E. Ferretti. A framework for multi-criteria argumentation-based decision making within a BDI agent. *JCS&T*, 14(1), 2014.
- [23] C. S. Toranzo, M. Errecalde, and E. Ferretti. Intention reconsideration like uncertain dichotomous choice model. In *XX CACIC*, 2014.
- [24] C. S. Toranzo, M. Errecalde, and E. Ferretti. On the use of agreement technologies for multi-criteria decision making within a BDI agent. In *IBERAMIA 2014*, pages 54–65, 2014.
- [25] M. Wooldridge. *An Introduction to MultiAgent Systems*. John Wiley & Sons, Chichester, England, 2002.
- [26] M. Wooldridge and N. R. Jennings. Intelligent agents: Theory and practice. *Knowledge Engineering Review*, 1995.

Introducir el Manejo de Múltiples Criterios de Comparación de Argumentos en Sistemas Argumentativos

Juan Carlos Teze^{1,2,3}, Sebastian Gottifredi^{1,2},
Alejandro Javier García^{1,2} y Guillermo Ricardo Simari^{1,2}

¹Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA)
Dep. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur (UNS),

²Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET),
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina

³Laboratorio de Investigación y Desarrollo en Inteligencia Artificial Concordia
Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos,
e-mail: {jct,sg,ajg,grs}@cs.uns.edu.ar

Resumen

Esta línea de investigación explora la incorporación del manejo de multiplicidad de criterios de comparación de argumentos en Sistemas Argumentativos. El objetivo general es mejorar las capacidades de razonamiento de estos sistemas introduciendo mecanismos para que puedan soportar varios criterios y elegir el que mejor se ajusta a las necesidades o preferencias del usuario. Como resultado se intentarán desarrollar herramientas concretas de interacción entre el usuario y el sistema, las cuales permitan especificar de una manera declarativa cual es el criterio que se debe utilizar como así también la información que este necesita para realizar su tarea.

Palabras clave: Razonamiento, Sistemas Argumentativos, Criterios de Comparación de Argumentos.

Contexto

Esta línea de investigación se llevará a cabo dentro del ámbito de colaboración entre el

Laboratorio de Investigación y Desarrollo (LIDIA) del Dep. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur; y el Laboratorio de Investigación y Desarrollo en Inteligencia Artificial Concordia (LIDIA Concordia) de la Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos. Está asociada a los siguientes proyectos de investigación:

- Argumentación y Dinámica de Creencias para mejorar las capacidades de razonamiento y representación de conocimiento en Sistemas Multi-agente. PGI 24/N035, financiados por la Universidad Nacional del Sur.

Además, este trabajo se realiza en el marco del desarrollo de una tesis doctoral para optar por el título de Doctor en Ciencias de la Computación del Departamento de Ciencias de la Computación de la Universidad Nacional del Sur.

Introducción

En los últimos años, la argumentación rebatible ha realizado un importante aporte a la Ciencias de la Computación, hecho que se refleja en el creciente número de aplicaciones del mundo real que la incluyen como formalización del razonamiento del sentido común. En este sentido, los sistemas argumentativos proponen una formalización de este tipo de razonamiento justamente utilizando como mecanismo de inferencia la argumentación rebatible. Intuitivamente, la argumentación rebatible provee formas de confrontar declaraciones contradictorias para determinar si alguna afirmación puede ser aceptada o rechazada. Para obtener una respuesta, el proceso de razonamiento argumentativo lleva a cabo una serie de etapas [15]. Una etapa muy importante es la comparación de argumentos en conflicto para decidir qué argumento prevalece; esto requiere introducir un criterio de comparación que haga frente a esta situación.

Distintos sistemas argumentativos adoptan diferentes criterios de comparación de argumentos. Uno de los criterios más populares en Inteligencia Artificial es el criterio de especificidad, que prefiere argumentos con información más específica. Sin embargo, existen diferentes posturas por parte de los investigadores respecto al criterio de comparación. Algunos investigadores [21, 12, 13], por ejemplo, consideran que especificidad no constituye un criterio general del razonamiento de sentido común, sino simplemente un criterio más que podría adoptarse. Otros investigadores sostienen que no existen principios generales, independientes del dominio, dado que éstos conllevan a un estado de indecisión en la mayoría de los casos, y que la información acerca del dominio constituye la herramienta más importante para decidir entre argumentos en conflicto [11, 21]. Por esta razón varios sistemas de argumentación se encuentran parametrizados respecto al criterio

de comparación, y se espera que se utilice el criterio que mejor se adecue al dominio que esta siendo representado. Por otra parte, otros investigadores [14] sostienen que los criterios son parte de la teoría de prueba, y por lo tanto sujetos a discusión, por lo que los sistemas argumentativos deberían permitir la construcción de argumentos acerca del criterio.

En esta línea de investigación proponemos mejorar las habilidades de razonamiento de los sistemas argumentativos incorporando herramientas programables que le permitan al usuario configurar de manera dinámica los criterios de comparación con los que el sistema cuenta para resolver los conflictos entre argumentos. Para lograr esto, se utilizará el sistema argumentativo llamado Defeasible Logic Programming (DeLP) [8], el cual es un formalismo que se caracteriza por tratar la comparación de argumentos de forma modular. Sin embargo, en DeLP una vez que se adopta un criterio (establecido en la configuración del sistema) acorde al dominio que se este representando, no existen aún en la literatura de argumentación formalismos que provean a este sistema de mecanismos para que se pueda seleccionar y cambiar dinámicamente el criterio de comparación de argumentos.

Líneas de investigación y desarrollo

El objetivo de esta línea de investigación es avanzar en la mejora de los sistemas argumentativos incorporando el manejo dinámico de criterio. Es decir, por un lado añadiendo mecanismos que permitan especificar de forma declarativa los criterios con los que dispone el sistema en un momento o situación dada, y por el otro, incorporando una herramienta que permita la selección de uno de estos criterios. Como resultado se buscarán desarrollar estas herramientas, considerando como mecanismo de

razonamiento un sistema argumentativo concreto como lo es DeLP. De este modo, se podría lograr un avance significativo para los sistemas argumentativos dentro del área de Inteligencia Artificial y Ciencias de la Computación.

Argumentación Rebatible

Argumentación es un mecanismo de razonamiento, el cual para determinar conclusiones se consideran los argumentos que las sustentan y los posibles conflictos entre ellos. En este tipo de razonamiento, todos los argumentos en conflicto son analizados para luego determinar qué conclusiones serán aceptadas. El propósito final de la argumentación es determinar las conclusiones aceptadas.

Desde hace tiempo, argumentación ha evolucionado como una atractiva propuesta para modelar el razonamiento de sentido común [6, 2, 3]. En particular, los sistemas argumentativos basados en reglas (SABR) son formalismos de argumentación basados en una lógica subyacente específica, la cual se emplea para expresar el conocimiento acerca del dominio que se intenta representar. Además, el conocimiento cuenta con un conjunto de reglas de inferencia lo cual permite construir argumentos a favor y en contra de una conclusión. Estos sistemas son de gran interés para la comunidad de Ciencias de la Computación dado que las reglas de inferencia permiten representar conocimiento de sentido común, posibilitando la construcción de argumentos de manera automática. Los SABR poseen características que los hacen especialmente aptos para su implementación computacional. Por otra parte, este tipo de sistemas es particularmente atractivo para la toma de decisiones y negociación [1, 4], y dentro del área de Inteligencia Artificial existe especial interés en abordar este tipo de problemas.

Programación en Lógica Rebatible (DeLP) y la Comparación entre Argumentos

Diversos trabajos en la literatura reconocen la importancia del uso de sistemas argumentativos como mecanismo de razonamiento en sistemas inteligentes [2, 7, 15]. En particular, esta línea de investigación está vinculada con un sistema argumentativo concreto, denominado Defeasible Logic Programming (DeLP) [8]. Este formalismo es un SABR que combina resultados de programación en lógica y argumentación rebatible, y ha sido aplicado exitosamente en diferentes dominios de aplicación (ver *e. g.* [5, 9, 10]). La Programación en Lógica Rebatible extiende a la Programación en Lógica permitiendo representar conocimiento contradictorio, mediante el uso de la negación fuerte, y conocimiento tentativo, incluyendo en la sintaxis un nuevo tipo de reglas: las reglas rebatibles.

Si bien al presentarse DeLP en [8] se asocia el criterio de Especificidad Generalizada desarrollado en [16] como criterio por defecto, en DeLP el criterio es un parámetro del sistema. De esta manera, al aplicar DeLP a un dominio concreto, puede emplearse el criterio que resulte adecuado para dicho dominio. Sin embargo, en la comunidad de argumentación no hay un consenso establecido acerca de qué criterio utilizar para evaluar argumentos. En consecuencia, incorporar mecanismos que le permitan al usuario tomar esta decisión programando la selección y configuración de criterios disponibles podría beneficiar y mejorar las capacidades de razonamiento de este tipo de sistemas en particular.

Resultados y Objetivos

Varios resultados obtenidos fueron publicados recientemente. A continuación se incluye un resumen de estas publicaciones.

Un primer resultado fue obtenido en el artículo *An Approach to Argumentative Reasoning Servers with Multiple Preference Criteria* [17]. En este formalismo se introduce un modelo de servidor recomendador basado en DeLP que provee recomendaciones a sus agentes clientes y la habilidad para que éstos puedan decidir cómo múltiples criterios de preferencia pueden ser combinados. Para que los clientes cuenten con esta capacidad varios operadores para combinar preferencias fueron propuestos. Por otra parte, el mecanismo de inferencia es quien resuelve las consultas utilizando los criterios indicados por el cliente. En este formalismo el cliente tiene la posibilidad de indicar en la consulta los criterios que debe utilizar el servidor y la forma en que debe combinarlos. Para lograr esto, la estructura de las consultas permitan incorporar una expresión que indique el criterio que el cliente desea que sea utilizado por el servidor para computar una respuesta. Por otro lado, en [19] se presentó una versión extendida del artículo introducido en [17]. En esta versión se incorporó una nueva sección que muestra la importancia del criterio de preferencia en el sistema DeLP como pieza clave en la construcción de los árboles de dialéctica.

En los artículos *An Approach to Argumentative Reasoning Servers with Conditions based Preference Criteria* [18] y *Improving argumentation-based recommender systems through context-adaptable selection criteria* [20] se desarrolla un nuevo enfoque para servidores recomendadores. En este enfoque el servidor puede utilizar un criterio diferente para responder cada consulta en particular. Para determinar el criterio que finalmente será usado por el servidor es necesario previamente evaluar una expresión condicional. El criterio resultante de evaluar dicha expresión va depender de si cierta información puede ser derivada o no del programa almacenado en el servidor. En pocas palabras, el modelo permite que el cliente pueda

especificar en la consulta a través de una condición el criterio que el servidor debe utilizar. Un cliente puede preferir cierta información sobre otra y asociar esta preferencia a criterios específicos. En particular, en [20] se realiza un análisis detallado de las expresiones condicionales y su semántica de evaluación considerando características importantes que conllevan a su optimización.

Por último, en la actualidad se está trabajando en el desarrollo de una implementación de un interprete DeLP que sea capaz de manipular de forma dinámica múltiples criterios de manera tal que se puedan incorporar al mismo los avances obtenidos hasta el momento.

Formación de Recursos Humanos

Los temas de esta línea de investigación están estrechamente relacionados con el desarrollo de la tesis doctoral del 1º autor del presente artículo.

Referencias

- [1] Amgoud, L., Parsons, S., Maudet, N.: Arguments, dialogue, and negotiation. a a 10(11), 02 (2000)
- [2] Bench-Capon, T.J.M., Dunne, P.E.: Argumentation in artificial intelligence. *Artif. Intell.* 171(10-15), 619–641 (2007)
- [3] Besnard, P., Hunter, A.: Elements of argumentation, vol. 47. MIT press Cambridge (2008)
- [4] Black, E., Hunter, A.: An inquiry dialogue system. *Autonomous Agents and Multi-Agent Systems* 19(2), 173–209 (2009)
- [5] Capobianco, M., Chesñevar, C.I., Simari, G.R.: Argumentation and the dyna-

- mics of warranted beliefs in changing environments. *Autonomous Agents and Multi-Agent Systems* 11(2), 127–151 (Sep 2005)
- [6] Chesñevar, C.I., Maguitman, A.G., Loui, R.P.: Logical models of argument. *ACM Computing Surveys* 32(4), 337–383 (2000)
- [7] Chesñevar, C.I., Maguitman, A.G., Simari, G.R.: Recommender system technologies based on argumentation 1. In: *Emerging Artificial Intelligence Applications in Computer Engineering*, pp. 50–73 (2007)
- [8] García, A.J., Simari, G.R.: Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming (TPLP)* 4(1-2), 95–138 (2004)
- [9] Gómez, S.A., Chesnevar, C.I., Simari, G.R.: Defeasible reasoning in web-based forms through argumentation. *International Journal of Information Technology & Decision Making* 7(01), 71–101 (2008)
- [10] Gottifredi, S., García, A.J., Simari, G.R.: Query-based argumentation in agent programming. In: *IBERAMIA*. pp. 284–295 (2010)
- [11] Konolige, K.: Defeasible argumentation in reasoning about events. In: *ISMIS*. pp. 380–390 (1988)
- [12] Pollock, J.L.: *Cognitive Carpentry: A Blueprint for How to Build a Person*. MIT Press (1995)
- [13] Prakken, H., Sartor, G.: A dialectical model of assessing conflicting arguments in legal reasoning. In: *Logical Models of Legal Argumentation*, pp. 175–211. Springer (1996)
- [14] Prakken, H., Sartor, G.: A system for defeasible argumentation, with defeasible priorities. In: *Practical Reasoning*, pp. 510–524. Springer (1996)
- [15] Rahwan, I., Simari, G.R.: *Argumentation in Artificial Intelligence*. Springer Publishing Company, Incorporated, 1st edn. (2009)
- [16] Stolzenburg, F., García, A.J., Chesñevar, C.I., Simari, G.R.: Computing generalized specificity. *Journal of Applied Non-Classical Logics* 13(1), 87–113 (2003)
- [17] Teze, J.C., Gottifredi, S., Garcia, A.J., Simari, G.R.: An approach to argumentative reasoning servers with multiple preference criteria. *XIV Simposio Argentino de Inteligencia Artificial* (2013)
- [18] Teze, J.C., Gottifredi, S., García, A.J., Simari, G.R.: An approach to argumentative reasoning servers with conditions based preference criteria. In: *XVIII Congreso Argentino de Ciencias de la Computación* (2013)
- [19] Teze, J.C., Gottifredi, S., García, A.J., Simari, G.R.: An approach to argumentative reasoning servers with multiple preference criteria. *Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial* 17(53), 68–78 (2014)
- [20] Teze, J.C., Gottifredi, S., García, A.J., Simari, G.R.: Improving argumentation-based recommender systems through context-adaptable selection criteria. *Expert Systems with Applications* 42(21), 8243–8258 (2015)
- [21] Vreeswijk, G.: *The feasibility of defeat in defeasible reasoning*. Springer (1993)

Definición de la infraestructura para procesos masivos de argumentación mediante aplicación de revisión de creencias y argumentación sobre ontologías Datalog+/-

Cristhian A. D. Deagustini María Vanina Martínez Marcelo A. Falappa
Guillermo R. Simari

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Departamento de Ciencias e Ingenierías de la Computación
Universidad Nacional del Sur
Alem 1253 - Bahía Blanca - Buenos Aires - Argentina
(0291) 459-5135

cadd@cs.uns.edu.ar, vanina.martinez@cs.ox.ac.uk, mfalappa@cs.uns.edu.ar, grs@cs.uns.edu.ar

Resumen

En los últimos tiempos, la colaboración y el intercambio de información se han vuelto aspectos cruciales de muchos sistemas. En estos entornos es de vital importancia definir métodos automáticos para resolver conflictos entre el conocimiento compartido por distintos sistemas. Este conocimiento es frecuentemente expresado a través de ontologías que pueden ser compartidas por los sistemas que utilizan el mismo.

En la presente investigación se busca la definición de métodos automáticos de resolución de conflictos en ontologías Datalog+/- . En base a lo logrado en este aspecto se buscará la adaptación del framework desarrollado para su aplicación tanto en la creación de federaciones de Bases de Datos (Data Federation) como en el intercambio de datos (Data Exchange). En estos campos de aplicación estos métodos podrán contribuir brindando la posibilidad de obtener de forma automática un esquema universal que respete tanto como sea posible a los originales manteniendo la coherencia del mismo con respecto a las restricciones de integridad impuestas a los datos, y definiendo que datos pueden ser mantenidos en la federación resolviendo incoherencias en el proceso.

Adicionalmente, se analizarán posibles extensiones a Datalog+/- basadas en formalismos de Argumentación Rebatible, teniendo en cuenta aspectos como la definición de relaciones de inferencia para estas ontologías aumentadas que tengan en cuenta los aspectos no-monótonos de la Argumentación Rebatible, o el impacto de tales relaciones en las conclusiones finales obtenidas y la complejidad de la obtención de las mismas.

Palabras Clave: Integración de Bases de Conocimiento, Revisión de Creencias, Representación de Conocimiento, Razonamiento, Argumentación Rebatible.

1. Contexto

Esta línea de investigación se lleva a cabo en el marco de los siguientes proyectos de investigación:

- **“Representación de conocimiento y razonamiento argumentativo: Herramientas inteligentes para la web y las bases de conocimiento”**. Director: Guillermo R. Simari. En evaluación, 01/01/15 – 31/12/2018. Unidad coordinadora: Universidad Nacional del Sur.
- **“Combinación de Revisión de Creencias y Argumentación para mejorar las capacidades de Razonamiento y modelado de la Dinámica de Conocimiento en Sistemas Multi-agente”**. Director: Marcelo A. Falappa. PIP 112-20110101000. Unidad coordinadora: Consejo Nacional de Investigaciones Científicas y Técnicas.

2. Introducción

En los últimos tiempos la integración e interacción entre diferentes sistemas se ha vuelto muy común, especialmente en entornos colaborativos introducidos desde el arribo de la Web Semántica [BLHL01], *e. g.*, e-commerce.

Sin embargo, la colaboración entre sistemas no siempre puede realizarse de manera directa. Hay ocasiones donde problemas de inconsistencia (e incoherencia) aparecen cuando tomamos el conocimiento provisto por diferentes entidades como uno solo. En formalismos de representación de conocimiento como las ontologías estas inconsistencias suelen aparecer como violaciones a las restricciones de integridad impuestas a los datos en las fuentes originales del conocimiento, que son violadas en el conocimiento integrado.

La resolución de incoherencias e inconsistencias en conocimiento es admitido como un problema importante

que debe ser atacado [GCS10, BQL07], especialmente en procesos de integración de conocimiento proveniente de fuentes diversas [BHP09, AK05]. En particular; el problema de incoherencia es conocido en formalismos de representación de conocimiento, especialmente en la comunidad de Lógicas Descriptivas (Description Logics - DL), donde ha sido analizado desde diferentes enfoques a través de los años [FHP⁺06, BB97]. En base a tales esfuerzos se han desarrollado adaptaciones del concepto para Datalog+/- [DMFS, DMFS15], que han sido de gran importancia para el desarrollo de métodos de integración que atiendan conflictos en el conocimiento ontológico además de aquél expresado en bases de datos.

La obtención de conocimiento coherente y consistente basado en distintas fuentes es un problema que aparece en muchos campos de la IA, *e. g.*, sistemas multi-agentes, y fue atacado desde diferentes enfoques. Un área que ha hecho muchos avances en la integración de fuentes inconsistentes de información es la Revisión de Creencias a través de la definición de procesos de Integración de Creencias (Belief Merging [BHA⁺01]). A través de los años, diferentes enfoques para la integración se han desarrollado, *e. g.*, [BKM91, Cho98, LS98, FKIRS12]. En un proceso de integración se toman varias Bases de Conocimiento (Knowledge Bases - KBs) que pueden ser consistentes en sí mismas pero generan conflicto cuando son consideradas juntas, y se obtiene una nueva KB que está libre de conflictos y refleja la información de las bases originales tanto como sea posible. Esta nueva base de conocimiento puede luego ser explotada mediante procesos de inferencia clásicos, los cuales no podrían ser aplicados directamente sobre la unión de las ontologías originales ya que los mismos no toleran inconsistencias.

Alternativamente, es posible tomar un camino diferente, y convivir con la inconsistencia, tratándola de manera local a las consultas que se intente resolver, en un proceso *ad-hoc*. Esto es, en lugar de resolver cada conflicto y limpiar la base de conocimiento para luego aplicar técnicas clásicas de inferencia, es posible *modificar cómo las inferencias son realizadas*, de forma que las respuestas dadas sean consistentes. Un área muy importante que maneja tal enfoque es la de Argumentación [RSvB09, GS04, GCS10].

Los métodos presentados suelen diferir en gran manera en la forma en que encarar los conflictos en el conocimiento; sin embargo, sin importar si se enfocan en eliminar los conflictos en la base de conocimiento o modificar cómo las respuestas son obtenidas para tolerar los conflictos sin resolverlos todos coinciden en que la estrategia debe ser de forma tal que la calidad de la información sea comprometida en la menor medida posible.

En esta línea de investigación encaramos la resolución de conflictos en ontologías a través de ambos enfoques. Esto es, utilizaremos tanto la consolidación e integración de ontologías desarrolladas en Datalog+/- [CGL12] como la definición de extensiones al lenguaje que permitan el uso de información rebatible [Pol87, SL92] a través de argumentación [RSvB09, GS04, GCS10], donde las técnicas a desarrollar se guiarán principalmente por este

concepto de preservación de la calidad de información, así como también por la eficiencia de los procesos utilizados; desarrollando técnicas que se adecuen a su uso en formalismos de representación de conocimientos aptos para el uso en entornos colaborativos como la Web Semántica. En los últimos años, las bases de conocimiento en forma de ontologías se han vuelto muy populares en estos entornos, ya que proveen formas de representar tanto los datos disponibles en sí como las restricciones impuestos a éstos. Además, el poder expresivo de las ontologías permite realizar tareas importantes en la integración de fuentes de conocimientos [Len02], y juega un rol preponderante en la Web Semántica[BLHL01].

Por lo tanto, es importante definir métodos que permitan el uso de ontologías resolviendo todos los problemas de incoherencia/inconsistencia que puedan aparecer. Esto podría ser el primer paso para la integración de otras fuentes de conocimiento que pueden ser expresadas mediante estas ontologías (*e. g.*, bases de datos relacionales). Es por esto que en esta línea de investigación nos enfocamos en la definición de la infraestructura necesaria para la realización de procesos masivos de argumentación, donde tales procesos se soportarán sobre ontologías desarrolladas en un lenguaje particular denominado Datalog+/- [CGL12]. La familia de lenguajes de ontologías Datalog permite un estilo modular de representación de conocimiento mediante el uso de reglas de forma similar a la usada en Programación Lógica, y su decidibilidad le permite manejar los volúmenes masivos de datos que podemos encontrar en aplicaciones hoy en día, haciéndola útil en diferentes campos como la consulta de ontologías, extracción de datos en web o intercambio de datos [LMS12]. En particular, la representación de conocimiento en ontologías Datalog+/- se lleva a cabo mediante el uso de (a) una Base de Datos: un conjunto de átomos que representan hechos acerca del mundo, *e. g.*, alumno(pedro) (b) Tuple-generating Dependencies - TGDs: reglas que nos permiten obtener nuevos átomos mediante la activación de las mismas como ser: alumno(X) \rightarrow persona(X), (c) Equality-generating Dependencies: reglas que restringen la generación de átomos, por ejemplo: doctor(D, P) \wedge doctor(D', P) \rightarrow D = D'; y (d) Negative Constraints - NCs: reglas que expresan relaciones que no pueden existir entre átomos, *e. g.*: alto(X) \wedge bajo(X) $\rightarrow \perp$. Los métodos a desarrollar se enfocarán, por lo tanto, en controlar la relación entre los átomos y las TGDs presentes en la ontología, de forma que se respeten las EGDs y NCs en la KB.

3. Líneas de Investigación y Desarrollo

Esta línea de investigación se enfoca en la definición de procesos de resolución de conflictos en ontologías Datalog+/- a través del uso de formalismos enfocados en la resolución de incoherencias e inconsistencias provenientes de las áreas de Revisión de Creencias y Argumentación. Para ello distintos ejes deben ser investigados,

que van desde la definición de incoherencias e inconsistencia en el entorno de ontologías Datalog+/- hasta las posibles aplicaciones que un método automático de integración de estas ontologías podría tener.

3.1. Definición de métodos de identificación de Incoherencias e Inconsistencias en Datalog+/-

Datalog+/- se ha vuelto un lenguaje muy popular en los últimos años, y numerosos estudios se han realizado acerca de sus propiedades de decibilidad y la complejidad asociada a la respuesta de consultas en estas ontologías. Sin embargo, no ha habido mucho estudio acerca de los aspectos de Representación de Conocimiento en Datalog+/. Muy poco trabajo se ha hecho acerca de inconsistencias en ontologías Datalog+/. Peor aún es la situación respecto del concepto de incoherencia (*i. e.*, la imposibilidad de satisfacer cierto conjunto de TGDs sin violar una NC).

Uno de los ejes de esta línea de investigación es la definición formal del concepto de incoherencia en Datalog+/-, tomando como partida esfuerzos similares que han sido realizados para otros formalismos de representación de conocimiento, principalmente Description Logics. Adicionalmente, se procederá a identificar las propiedades que llevan a que un conjunto de TGDs sea incoherente, y las que hacen que una ontología Datalog+/- se vuelva inconsistente. De esta manera se podrán identificar tales casos, lo que será el primer paso para la posterior resolución de tales problemas.

3.2. Resolución de Incoherencias e Inconsistencias en Datalog+/-

Una vez que se tiene definidos e identificados los conjuntos incoherentes de TGDs y aquellos conjuntos de átomos que provocan inconsistencias en la unión de varias ontologías Datalog+/-, se debe proceder a la resolución de estos conflictos. En esta línea de investigación esto se hará mediante la aplicación de técnicas derivadas de la Revisión de Creencias denominadas Kernel Contraction.

Este tipo de técnicas resuelve conflictos de incoherencia/inconsistencia tomando los conjuntos conflictivos mínimos y eligiendo de alguna forma que elemento remover de los mismos para solucionar el problema. En el caso de integración de ontologías Datalog+/- esto es la remoción de ciertos átomos y ciertas TGDs de la unión de todas las ontologías que se está integrando. Adicionalmente, se puede pensar en la definición de métodos de debilitamiento de reglas, en lugar de la remoción de las mismas. Esto no es una tarea trivial, ya que hay muchos aspectos a definir, por ejemplo como elegir el mejor candidato entre los átomos o TGDs que pueden eliminarse, lo que a su vez lleva a definir formas (automáticas) de obtener órdenes entre los candidatos. Para esto se procederá a definir operadores de integración de ontologías Datalog+/-, así como se darán las propiedades esperadas de tales operadores y se definirán métodos de obtención

de operadores de tales características mediante Teoremas de Representación.

3.3. Posibles extensiones a Datalog+/- mediante Argumentación Rebatible

Otro aspecto a considerar dentro de la línea de investigación es como se beneficiaría Datalog+/- de otros formalismos de representación de conocimientos con propiedades diferentes a aquellas presentes en Datalog+/. Particularmente, proponemos analizar posibles extensiones a las ontologías Datalog+/- basadas en el uso de formalismos de Argumentación Rebatible como Programación Lógica Rebatible (Defeasible Logic Programming - DeLP) [GS04].

Para tales extensiones se deberán analizar diferentes aspectos. Por ejemplo, la relación de inferencia en estas ontologías aumentadas deberá tener en cuenta los aspectos no-monótonos de la Argumentación Rebatible, llevando a que se modifique la forma en que una consulta es respondida respecto del proceso actual en Datalog+/-, ya que se deberá tener en cuenta el análisis dialéctico llevado a cabo por DeLP antes de responder la misma.

Estas modificaciones en la forma en que información es inferida traerá aparejado un impacto en las inferencias finales de las ontologías en aquellos casos en donde las ontologías no son coherentes o consistentes, proveyendo otra forma de integrar las mismas (*i. e.*, simplemente considerarlas juntas sin importar problemas de incoherencia e inconsistencia, y dejar que el proceso argumentativo los resuelva).

3.4. Integración de otras fuentes de datos a través del Merging de Ontologías Datalog+/-

Finalmente, se analizará nuestro framework de integración de ontologías Datalog+/- como medio de integración de otras fuentes de datos. Principalmente nos enfocaremos en la integración automática de Bases de Datos Relacionales.

Para esto, primeramente definiremos métodos para expresar bases de datos relacionales a través de ontologías Datalog+/-, tanto los datos en sí como aspectos relacionados al esquema de las mismas, *e. g.*, las dependencias funcionales.

Una vez logrado esto se podría utilizar los métodos de integración de ontologías Datalog+/- para obtener una federación de las bases de datos expresadas, ya que la ontología final resultante de la integración brindaría un esquema integrador de las mismas así como los datos que serían parte de la federación, manteniendo a su vez la coherencia de las restricciones de integridad respecto del esquema unificado y la consistencia de los datos almacenados respecto de las dependencias funcionales.

4. Resultados y Objetivos

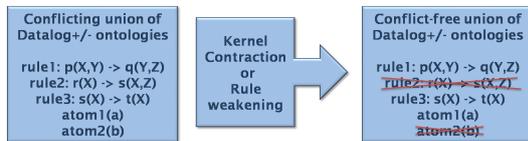
El objetivo general de este trabajo de investigación es el diseño y construcción de la infraestructura necesaria para la realización de procesos de argumentación en una escala masiva valiéndose de datos almacenados en Bases de Datos heterogéneas integradas en Federaciones.

En el presente trabajo se propone la representación de tales bases de datos a través de la utilización de ontologías Datalog+/-, las cuales poseen una expresividad y una tratabilidad adecuadas para tal fin. De esta forma, todos los desarrollos que se logren en Datalog+/- serían transferibles a bases de datos (y otras tecnologías estrictamente menos expresivas que Datalog+/- como ciertas Description Logics, *e. g.*, DL-Lite o \mathcal{EL}).

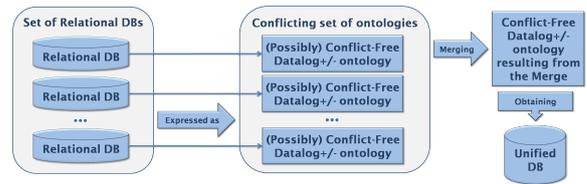
Hay dos aspectos principales a tener en cuenta para la consecución del objetivo planteado de realización de procesos masivos de argumentación. Por un lado tenemos la definición de métodos automáticos que permitan la creación de vistas unificadas de varias bases de datos como forma de construir las federaciones de bases de datos, de forma que los procesos de argumentación accedan a las mismas para dar soporte a los argumentos. Por el otro, es posible modificar la forma en que la información es inferida en Datalog+/-, de forma que la misma sea tolerante a inconsistencias e incoherencias. Debido a esta dualidad respecto a la consecución del objetivo general se plantean objetivos específicos que abarcan dos temáticas diferentes aunque relacionadas.

Respecto de la definición de métodos automáticos como forma de construir federaciones:

- Definición de métodos de consolidación de ontologías Datalog+/- que permitan el manejo de inconsistencia e incoherencia aprovechando tanto enfoques ya clásicos en la literatura de revisión de creencias como Kernel Contraction [Han94, Han99] como así también refinamientos desarrollados en la presente línea de investigación (Cluster Contraction [DMFS14b, DMFS14a]).



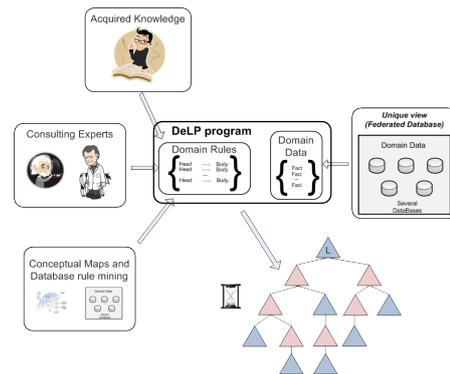
- En base a los métodos desarrollados procederemos a definir métodos que permitan la integración de ontologías Datalog+/-, mediante la utilización de la consolidación de la unión de ontologías Datalog+/- . Una vez logrado esto se podría utilizar los métodos de integración para obtener una federación de las bases de datos expresadas, ya que la ontología final resultante de la integración brindaría un esquema integrador de las mismas así como los datos que serían parte de la federación, manteniendo a su vez la coherencia de las restricciones de integridad respecto del esquema unificado y la consistencia de los datos almacenados respecto de las dependencias funcionales, como se muestra a continuación.



A su vez, respecto de la modificación de las relaciones de inferencia utilizadas en Datalog+/- para hacerlas tolerantes a inconsistencia e incoherencia tenemos:

- Desarrollo y análisis de posibles extensiones a la familia de lenguajes Datalog+/- basadas en el uso de formalismos de Argumentación Rebatible como Programación Lógica Rebatible (Defeasible Logic Programming - DeLP) [GS04], avanzando sobre los resultados introducidos por esta línea de investigación [MDFS14, DMFS15]. Estos avances se han realizado sobre conjuntos restringidos de ontologías Datalog+/- . Como continuación de tales trabajos se procederá a generalizar tales desarrollos para la totalidad del conjunto de ontologías Datalog+/- .
- Análisis del incremento que tales desarrollos provoquen en la cantidad de información que puede ser inferida de las ontologías Datalog+/- (un primer avance al respecto puede verse en [DMFS15]) como así también del trade-off que el uso de argumentación en Datalog+/- induce respecto del aumento de inferencias y la complejidad y eficiencia en la resolución de consultas.

Finalmente, se pretende analizar el uso de los métodos desarrollados en distintas aplicaciones como sistemas de soporte de decisiones (Decision Support Systems - DSS) y sistemas de recomendación (Recommender Systems - RS). Un esquema general que tales aplicaciones pueden seguir es mostrado a continuación.



5. Formación de Recursos Humanos

En la presente línea de investigación se enmarca el desarrollo de una tesis de posgrado en el Doctorado en Ciencias de la Computación del Departamento de Ciencias e Ingenierías de la Computación de la Universidad Nacional del Sur.

Referencias

- [AK05] L. Amgoud and S. Kaci. An argumentation framework for merging conflicting knowledge bases: The prioritized case. In *ECSQA-RU 2005*, pages 527–538, 2005.
- [BB97] D. Beneventano and S. Bergamaschi. Incoherence and subsumption for recursive views and queries in object-oriented data models. *Data Knowl. Eng.*, 21(3):217–252, 1997.
- [BHA⁺01] I. Bloch, A. Hunter, A. Appriou, A. Ayoun, S. Benferhat, P. Besnard, L. Cholvy, R. M. Cooke, F. Cuppens, D. Dubois, H. Fargier, M. Grabisch, R. Kruse, J. Lang, S. Moral, H. Prade, A. Saffiotti, P. Smets, and C. Sossai. Fusion: General concepts and characteristics. *Int. J. Intell. Syst.*, 16(10):1107–1134, 2001.
- [BHP09] E. Black, A. Hunter, and J. Z. Pan. An argument-based approach to using multiple ontologies. In *SUM*, pages 68–79, 2009.
- [BKM91] C. Baral, S. Kraus, and J. Minker. Combining multiple knowledge bases. *IEEE Trans. Knowl. Data Eng.*, 3(2):208–220, 1991.
- [BLHL01] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):3443, 2001.
- [BQL07] D. A. Bell, G. Qi, and W. Liu. Approaches to inconsistency handling in description-logic based ontologies. In *OTM Workshops (2)*, pages 1303–1311, 2007.
- [CGL12] A. Cali, G. Gottlob, and T. Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *J. Web Sem.*, 14:57–83, 2012.
- [Cho98] Laurence Cholvy. Reasoning about merged information. In *Belief Change*, volume 3, pages 233–263. Springer Netherlands, 1998.
- [DMFS] C. A. D. Deagustini, M. V. Martinez, M. A. Falappa, and G. R. Simari. Datalog+/- ontology consolidation. *J. Artif. Intell. Res. (JAIR)*. Under Review.
- [DMFS14a] C. A. D. Deagustini, M. V. Martinez, M. A. Falappa, and G. R. Simari. Improving inconsistency resolution by considering global conflicts. In *SUM*, pages 120–133, 2014.
- [DMFS14b] C. A. D. Deagustini, M. V. Martinez, M. A. Falappa, and G. R. Simari. Inconsistency resolution and global conflicts. In *ECAI*, pages 991–992, 2014.
- [DMFS15] C. A. D. Deagustini, M. V. Martinez, M. A. Falappa, and G. R. Simari. On the influence of incoherence in inconsistency-tolerant semantics for datalog+/- . In *ONTOLP '15, IJCAI-2015, Buenos Aires, Argentina, Julio, 2015, Proceedings*, 2015.
- [FHP⁺06] G. Flouris, Z. Huang, J. Z. Pan, D. Plexousakis, and H. Wache. Inconsistencies, negations and changes in ontologies. In *AAAI*, pages 1295–1300. AAAI Press, 2006.
- [FKIRS12] M. A. Falappa, G. Kern-Isberner, M. Reis, and G. R. Simari. Prioritized and non-prioritized multiple change on belief bases. *J. of Philos. Logic*, 41(1):77–113, 2012.
- [GCS10] S. A. Gómez, C. I. Chesñevar, and G. R. Simari. Reasoning with inconsistent ontologies through argumentation. *Appl. Artif. Intell.*, 24(1&2):102–148, 2010.
- [GS04] A. J. García and G. R. Simari. Defeasible logic programming: An argumentative approach. *TPLP*, 4(1-2):95–138, 2004.
- [Han94] S. O. Hansson. Kernel contraction. *J. Symb. Log.*, 59(3):845–859, 1994.
- [Han99] S. O. Hansson. *A Textbook of Belief Dynamics*. Kluwer Academic Publishers, 1999.
- [Len02] M. Lenzerini. Data integration: A theoretical perspective. In *PODS*, pages 233–246, 2002.
- [LMS12] T. Lukasiewicz, M. V. Martinez, and G. I. Simari. Inconsistency handling in datalog+/- ontologies. In *Proc. of ECAI*, pages 558–563, 2012.
- [LS98] P. Liberatore and M. Schaerf. Arbitration (or how to merge knowledge bases). *IEEE Trans. Knowl. Data Eng.*, 10(1):76–90, 1998.
- [MDFS14] M. V. Martinez, C. A. D. Deagustini, M. A. Falappa, and G. R. Simari. Inconsistency-tolerant reasoning in datalog+/- ontologies via an argumentative semantics. In *IBERAMIA '14*, pages 15–27, 2014.
- [Pol87] John L. Pollock. Defeasible reasoning. *Cognitive Science*, 11(4):481–518, 1987.
- [RSvB09] I. Rahwan, G. R. Simari, and J. van Benthem. *Argumentation in artificial intelligence*, volume 47. Springer, 2009.
- [SL92] G. R. Simari and R. P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artif. Intell.*, 53(2-3):125–157, 1992.

Razonamiento basado en la tolerancia a la inconsistencia

Mariana Virginia Etcheber María Vanina Martínez
Guillermo R. Simari

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Departamento de Ciencias e Ingenierías de la Computación
Universidad Nacional del Sur
Alem 1253 - Bahía Blanca - Buenos Aires - Argentina
(0291) 459-5135
mve@cs.uns.edu.ar, mvm@cs.uns.edu.ar, grs@cs.uns.edu.ar

Resumen

En los últimos 20 años se ha dado un gran incremento en la proliferación masiva de datos originado por la aplicación de nuevas tecnologías. Estas cantidades de información se almacenan en grandes repositorios de datos que quedan disponibles para su uso cotidiano como bases de conocimiento (KB), por ejemplo para utilizar como soporte en la toma de decisiones. En este contexto la presencia de información conflictiva (inconsistencia) es ubicua; dada la cantidad y variedad de procesos y aplicaciones que modelan situaciones del mundo real, las actualizaciones de los mismos producen inevitablemente inconsistencias que se almacenan en las KB. Por lo tanto, la inconsistencia es ineludible y esperable en KBs del mundo real. Dependiendo del dominio de aplicación pueden aparecer inconsistencias que necesitan ser resueltas pero la decisión de resolverlas, el enfoque de la resolución y cuando deben ser resueltas deberían ser procesos sensibles al contexto. Por ejemplo, ciertos tipos de conflictos en los datos puede ser visto como aceptable y hasta deseable en un sistema con razonamiento utilizado para detectar fraudes; en otros escenarios la resolución inmediata de inconsistencias puede resultar en pérdida de información valiosa. Esta línea de investigación propone explorar el área de razonamiento basado en tolerancia a la inconsistencia en Inteligencia Artificial y en Teoría de Base de Datos. Esta tarea consiste en realizar la recolección, análisis y crítica de los diferentes métodos propuestos en la literatura sobre el manejo de inconsistencias en el proceso de razonamiento y respuesta a consultas y su consolidación en un trabajo que relacione estas perspectivas diferentes.

Palabras Clave: Representación de conocimiento y razonamiento, Razonamiento tolerante la inconsistencia, Bases de datos, Lenguajes ontológicos

1. Contexto

La presente línea de investigación se encuentra inserta en el marco del proyecto PRH-2014-0007 “*Respuesta*

de consultas a bases de conocimiento sociales con incertidumbre utilizando las preferencias de los usuarios”, coordinado por la Agencia Nacional de Promoción Científica y Tecnológica, Ministerio de Ciencias, Tecnología e Innovación Productiva, y llevado a cabo dentro del Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur.

2. Introducción

Esta investigación se centra en el estudio de métodos y arquitecturas para razonamiento tolerante a la inconsistencia. En las últimas dos décadas se han desarrollado en diferentes sub-áreas de la Inteligencia Artificial y Teoría de Bases de Datos, varias propuestas al respecto enfocadas a resolver problemas de distinta naturaleza pero que comparten el mismo espíritu de conservar la mayor cantidad de información posible en presencia de potenciales conflictos. Sin embargo, las propuestas no son completamente satisfactorias en relación al compromiso en la calidad del resultado y la eficiencia computacional del proceso de razonamiento realizado bajo estas condiciones; por lo tanto, se considera que el problema de investigación aún está abierto. La influencia de la informática en las comunicaciones ha producido una situación en la que la sociedad se nutre de información. La mayor parte de la información está en forma “cruda”, no elaborada como datos elementales. La tecnología informática disponible, actuando sobre el espacio comunicacional, permite capturar, administrar y almacenar grandes cantidades de datos en repositorios masivos que quedan disponibles para su uso cotidiano como base de conocimiento. Estos repositorios alimentan una gran variedad de sistemas de computación, tales como los que brindan soporte para la toma de decisiones, sistemas de comando para control de emergencias, gestión de desastres, aplicaciones en la web semántica, entre muchos otros. En este contexto, la inconsistencia en las bases de conocimiento es ubicua. Esto se explica a través de la cantidad y variedad de procesos y aplicaciones que modelan situaciones del mundo real, que inevitablemente produce in-

formación conflictiva durante las actualizaciones y que es almacenada en las bases de conocimiento. La inconsistencia en la información es ineludible y esperable en las bases de conocimiento del mundo real, y como tal, su existencia reconocida en los sistemas de información con los que interactuamos cotidianamente y su manejo debería ser formalizado. Cuanto más grande es una base de conocimiento, menos posibilidades existen que los datos contenidos en ella sean completos y coherentes (o consistentes). Durante mucho tiempo, en diferentes áreas de Ciencias de la Computación, se sostuvo la postura general de que las bases de datos, bases de conocimiento, y especificaciones de software, debían estar completamente libres de inconsistencia; posteriormente, el problema se abordó formalmente aislando o resolviendo las inconsistencias localmente. Desde el punto de vista del sentido común, un conjunto de datos que contenga la forma α y $\neg\alpha$ no debería existir en una base de conocimiento. De existir el conflicto, este debería resolverse. Sin embargo, hay situaciones en las que pueden aceptarse su almacenamiento conjunto, dado que no debería entorpecer el razonamiento sobre el resto de la información que mantiene su consistencia. También es posible considerar la inconsistencia como conocimiento extra que permite identificar situaciones anómalas y razonar adecuadamente con esa información. Consideremos, por ejemplo, una base de datos del impuesto sobre la renta donde la información contradictoria sobre un contribuyente puede ser una prueba útil en una investigación de fraude. Tal vez el contribuyente haya completado un formulario que indique que tiene seis hijos (de ahí los beneficios fiscales para eso) y completado otro que establece que no tiene ninguno. Aquí, esta información contradictoria necesita ser almacenada, descubierta y utilizada apropiadamente para resolver la situación. La inconsistencia puede ser útil para dirigir el razonamiento y la investigación de procesos naturales de argumentación, búsqueda de información, interacción multi-agente, adquisición y perfeccionamiento, adaptación y aprendizaje de conocimiento. En un sentido, la inconsistencia puede ser vista como aceptable y hasta deseable en un sistema, siempre y cuando el sistema tenga los mecanismos apropiados para actuar sobre ellas, esto es, el razonamiento que se utiliza es supra-clásico. Otro conjunto de problemas de gran interés es la cuestión de cuándo resolver las inconsistencias. La resolución inmediata de inconsistencias puede resultar en la pérdida de información valiosa si se elige arbitrariamente sobre lo que debe rechazar. Consideremos por ejemplo la etapa de ingeniería de software de captura de requerimientos. Aquí una resolución prematura puede forzar una decisión arbitraria hecha sin que la elección sea ponderada adecuadamente. Esta forma de proceder puede limitar excesivamente la captura de requisitos del proceso. Por supuesto, dependiendo del dominio de aplicación, pueden aparecer inconsistencias que necesitan ser resueltas, pero la decisión de resolverlas, el enfoque de la resolución y cuando deben ser resueltas deberían ser procesos sensibles al contexto. En los últimos veinte años, a raíz de la aplicación de las

nuevas tecnologías de información se ha producido la proliferación masiva de datos en la ciencia básica (Big Data) lo que dificulta a los científicos mantenerse al día con los resultados y tecnologías desarrolladas. Esto lleva, por ejemplo, a que los científicos malgasten tiempo, esfuerzo, recursos y talento. Los grandes repositorios de datos de la experiencia están disponibles a costos bajos e incluso gratuitos, listos para ser explotados por procesos informacionales. Por esta razón, este plan de trabajo propone una investigación exploratoria del área de razonamiento basado en tolerancia a la inconsistencia en Inteligencia Artificial y en Teoría de Base de Datos. Esta tarea implica realizar la recolección, análisis y crítica de los diferentes métodos propuestos en la literatura sobre el manejo de inconsistencias en el proceso de razonamiento y respuesta a consultas y su consolidación en un trabajo que relacione estas perspectivas diferentes.

3. Líneas de Investigación y Desarrollo

Intuitivamente, una base de conocimiento es inconsistente cuando el proceso de inferencia que opera sobre ella obtiene conclusiones conflictivas. La gravedad de esta situación depende de la infraestructura inferencial disponible. En el caso de la deducción en lógica clásica (proposicional o de primer orden), la maquinaria deductiva es capaz de deducir cualquier sentencia del lenguaje. Este problema es descrito como el principio de explosión o de trivialización (*ex falso quodlibet* en Latín), que resulta claramente en una situación indeseable. Un sistema que usa el lenguaje de la lógica de primer orden para representar el conocimiento almacenado en la KB, puede tener diferentes respuestas de acuerdo la máquina inferencial que se elija. Si se sigue una aproximación deductiva clásica, una KB conteniendo inconsistencia sufrirá el problema del *ex falso quodlibet*, pero podría evitarlo si su máquina de inferencia está construida modelando alguna forma de razonamiento rebatible cuya misma definición contempla este problema resolviéndolo por medio del análisis de las consecuencias de aceptar alguna de las alternativas considerando como se llega a las conclusiones. La tolerancia a la inconsistencia trata sobre la definición de estos procesos inferenciales [6]. La presente línea de investigación se enfoca en el estudio de las diversas (a veces aparentemente distintas) maneras de hacer frente a la presencia de información inconsistente o conflictiva. Describiremos brevemente algunos enfoques “clásicos” de abordar esta problemática.

Chequeo y reparación de inconsistencia. En bases de datos (DB), la inconsistencia es una noción relativa a la satisfacción de un determinado conjunto de restricciones de integridad (CI), estas restricciones de integridad establecen propiedades de los estados de una DB admisible; imponen restricciones semánticas sobre los datos con el fin de capturar su correspondencia con lo que está siendo modelado en la DB. Decimos que una DB (que es posible modelar como una estructura de lógica de primer

orden [27] es inconsistente cuando las restricciones de integridad, expresadas como fórmulas lógicas, no están satisfechas por la DB. De esta manera, el chequeo de satisfacción de las restricciones de integridad se puede hacer fácilmente planteando y respondiendo una consulta desde/a la DB. Si bien este es un proceso simple, la dinámica y evolución en el contenido de las DBs hacen necesario realizar este chequeo periódicamente. A partir del trabajo de Arenas *et al.* [3] el área de “reparación y respuesta consistente” (CQA Consistent Query Answering en Inglés) a consultas en bases de datos relacionales ha ganado mucha atención ya que, en ese trabajo se define la construcción de un modelo teórico para una reparación de una DB: un reparación de una DB inconsistente es un modelo del conjunto de restricciones de integridad (ICs) que es minimal con respecto a la DB original, *i.e.*, “tan cercano a la DB original como sea posible”. Las reparaciones a realizar a una DB pueden no ser únicas, y en el caso general pueden existir muchas de ellas. Una semántica ampliamente aceptada para consultar una DB posiblemente inconsistente es la de “respuestas consistentes”. Una respuesta consistente para una consulta sobre una DB posiblemente inconsistente es el conjunto de tuplas que aparecen en la respuesta a la consulta para cada posible reparación. CQA impone consistencia en el tiempo de consulta como una alternativa al requerimiento a nivel de instancia como lo hacen las técnicas convencionales de limpieza de datos. Esto permite centrarnos en una pequeña parte de la base de datos para la cual las reparaciones no necesitan ser computadas o pueden ser computadas más fácilmente. En [12] se presenta un marco formal para reparación de DBs basado en costos, el cual permite encontrar “buenas” reparaciones para bases de datos que exhiben inconsistencia en forma de violación de dependencias funcionales o de inclusión. Claramente, CQA propone un enfoque cauteloso, y puede conducir a una gran pérdida de información. Aunque existen varios trabajos sobre el problema de reparación y consulta inconsistente de datos considerando diferentes clases de consultas y restricciones, solo recientemente se han desarrollado propuestas que centran la atención en mejorar la calidad de las respuestas [10, 29]; estos enfoques se centran en el desarrollo de estrategias de reparación específicas y en mejorar su calidad de acuerdo a algún criterio especificado por el administrador de base de datos de acuerdo con las preferencias del usuario.

Lógicas Paraconsistentes. Dados los inconvenientes que genera la inconsistencia en la lógica clásica, es evidente que el razonamiento con inconsistencia implica un compromiso con la maquinaria de inferencia de la lógica que se utiliza para modelar el conocimiento. Algunos de los tipos de lógica paraconsistente que han sido usados en sistemas computacionales inteligentes incluyen:

- Lógicas Débilmente Negativas, utilizan el lenguaje clásico completo para un subconjunto de la teoría clásica de prueba [13],
- Lógicas de Cuatro Valores, utilizan subconjuntos del lenguaje clásico y subconjuntos de la teoría clásica de prueba y semántica de cuatro valores [4],

- Sistemas Signados que implican renombrar todos los literales en una teoría y luego restaurar algo de la teoría original de forma progresiva añadiendo equivalencias formales entre los literales originales y sus cambios de nombre [8].
- Lógicas Cuasi-Clásicas que utilizan la teoría de prueba clásica pero restringen la noción de prueba de deducción natural [18].

Cada una de las propuestas tiene ventajas y desventajas, y en general se comportan de manera muy diferente. Seleccionar una lógica paraconsistente apropiada para una aplicación depende de los requisitos de esta aplicación y ninguno puede ser considerado como perfecto para el manejo de información inconsistente en general.

Sistemas de Argumentación. Argumentar es un proceso natural que los seres humanos usan para decidir cuáles son sus creencias. Al argumentar se comparan razones (en la forma de argumentos) en favor y en contra de aceptar una afirmación. Un argumento estructuralmente sostiene una afirmación, y el sostén se obtiene siguiendo alguna forma de razonamiento a partir de un conjunto de premisas. Un contraargumento es un argumento que ataca la conclusión, las premisas o la conexión entre las premisas y la afirmación que representa la conclusión. Argumentación es dar razones para justificar afirmaciones e involucra una audiencia que decide si la afirmación se acepta o no, todo esto en el contexto de un desacuerdo. Esta forma de razonamiento ha encontrado un lugar importante en el estudio de los procesos asociados con el pensamiento de sentido común y también en el modelamiento computacional de muchas actividades humanas. En particular, la aplicación del análisis dialéctico del contenido de bases de conocimiento posiblemente inconsistentes ha dado lugar un número importante de propuestas, entre las que podemos mencionar las siguientes:

- Sistemas basados en la formación de coaliciones: Se basan en la identificación de grupos de argumentos que se defienden unos a otros contra contraargumentos al unirse para defensa propia [15].
- Sistemas basados en coherencia: La estrategia es razonar con subconjuntos consistentes de la base de conocimiento. Esto está estrechamente relacionado con un enfoque de eliminación de información de la base de conocimiento que está causando inconsistencia [24, 11, 5, 2, 7].
- Sistemas basados en lógica rebatible: Existen diferentes propuestas para considerar la rebatibilidad en argumentación; la característica común es la incorporación de una implicación rebatible en un lenguaje lógico (puede ser clásico o de programación en lógica). A partir de la conceptualización de Pollock [26] de las nociones básicas del proceso de argumentación tales como la noción de razones, razones prima facie, derrotadores, refutadores de derrotadores y subvaloración de derrotadores, en términos de una lógica formal, surgieron otras propuestas como las de argumentos abstractos [30], lógica

condicional [25], programación en lógica rebatible (DeLP) [16], ASPIC framework [9], entre otras.

Medidas de Inconsistencia. Dado una KB inconsistente, es posible que necesitemos saber más sobre la naturaleza de esa inconsistencia y de la información que ofrece la KB. En cierto sentido, podemos desear hacer el análisis de inconsistencia en base a nociones que puedan ser medidas. El trabajo seminal sobre la medición de inconsistencia es de Shannon [28], el cual se basa en la teoría de probabilidades y puede ser utilizado en un entorno lógico cuando los mundos son los eventos posibles. Este trabajo es también la base de la obra de Lozinskii [23] para la definición de cantidad de información de una fórmula (o KB) en lógica proposicional. Pero esta definición no es adecuada cuando la base de conocimiento es inconsistente. En este caso, la base de conocimiento no posee un modelo clásico, por lo que no existen “eventos” para contar. Para hacer frente a esto, se consideran modelos de subconjuntos consistentes maximales de la base de conocimiento. Alternativamente, se puede medir la contradicción contenida en una base de conocimiento. En la lógica clásica se utiliza una medida binaria de contradicción: consistente o inconsistente; claramente, esta medida carece de sentido en la presencia de inconsistencia. Sin embargo, dado que en la actualidad hay una serie de lógicas paraconsistentes desarrolladas para obtener conclusiones no triviales de una base de conocimiento inconsistente, es necesaria la utilización de medidas más finas. Algunas propuestas interesantes incluyen: análisis basado en consistencia que se centraliza en los subconjuntos consistentes e inconsistentes de la base de conocimiento [19], análisis teórico de información que se adapta la medida de información de Shannon [23, 31], análisis semánticos probabilísticos que consideran distribuciones de probabilidad consistente maximal sobre un conjunto de fórmulas [21] entre otros.

Revisión de Creencias. Dada una base de conocimiento KB y una fórmula p de revisión, la teoría de revisión de creencias se refiere a las propiedades que debe contener una noción racional para la actualización con respecto a la fórmula p . Si $KB \cup \{p\}$ es inconsistente, la teoría de revisión de creencias asume el requisito que el conocimiento debe ser revisado de manera que la base de conocimiento resultante sea consistente. Los axiomas del sistema AGM [1, 17] delimitan el comportamiento de las funciones de revisión de conjuntos de creencias (conjunto de todas las inferencias obtenidas a partir de un conjunto de fórmulas). Un operador de revisión intenta cambiar lo menos posible del conjunto de creencias con el fin de incluir la nueva información. Este requerimiento de cambiar lo menos posible se opone al cambio de un conjunto consistente a uno inconsistente, i.e., algunas creencias serán eliminadas con el fin de mantener la consistencia. Obtener sistemas computacionalmente eficientes y efectivos que cumplan con los postulados de AGM ha resultado ser un reto. Ha habido muchos desarrollos de teoría de revisión de creencias como iteración de revisión de creencias [14, 22] y revisión de creencias para actualización de base de datos [20].

4. Resultados y Objetivos

En esta investigación se estudiarán los diferentes formalismos para el razonamiento en presencia de conocimiento inconsistente que se han desarrollado en las áreas de Inteligencia Artificial y Teoría de Bases de Datos en las últimas dos décadas. El objetivo principal de este plan de trabajo es generar un panorama global del estado del arte en esta área que permitirá, por un lado, clarificar el grado de avance en el área y, por otro, identificar posibles caminos de investigación a seguir con el objetivo de desarrollar formalismos que manejen la información inconsistente de una manera adecuada para la utilización en sistemas verdaderamente inteligentes y útiles para los seres humanos. Para este desarrollo se han fijado varios objetivos parciales tales como:

- Se estudiarán las diferentes propuestas desde un punto de vista formal que demuestre claramente las bondades y/o falencias de cada propuesta.
- Se clasificarán y compararán los distintos formalismos; esta clasificación permitirá realizar una comparación formal entre los métodos, como así también identificar nuevas direcciones de investigación y desarrollo en pos de la generación de sistemas inteligentes con capacidades de razonamiento más cercanas a la de los seres humanos que permitan un mejor desempeño de éstos últimos en diversas tareas computacionales.

Los objetivos listados tienen como fin clarificar el grado de avance en el área y, por otro lado, identificar nuevas vertientes de investigación. Los resultados esperados pueden resumirse de la siguiente manera:

- Estudio de los formalismos existentes en las siguientes áreas: (1) Lógicas paraconsistentes, (2) Reparación y respuestas consistentes a consultas en bases de datos relacionales y lenguajes ontológicos, (3) Sistemas argumentativos y (4) Revisión de creencias.
- Desarrollo de ejemplos que muestren pros y contra para cada formalismo.
- Definición de un conjunto de propiedades deseables para el razonamiento basado en la tolerancia a la inconsistencia para la clasificación de las diferentes propuestas dependiendo de si cumplen con dichas propiedades (o hasta qué punto lo hacen) o no.
- Clasificación y comparación de los diferentes formalismos en base al conjunto de propiedades.

5. Formación de Recursos Humanos

En la presente línea de investigación se enmarca el desarrollo de una tesis de posgrado en el Master en Ciencias de la Computación del Departamento de Ciencias e Ingenierías de la Computación de la Universidad Nacional del Sur.

Referencias

- [1] ALCHOURRÓN, C. E., GÄRDENFORS, P., AND MAKINSON, D. On the logic of theory change: Partial meet contraction and revision functions. *J. Symb. Log.* 50, 2 (1985), 510–530.
- [2] AMGOUD, L., AND CAYROL, C. On the acceptability of arguments in preference-based argumentation. In *Proc. of UAI* (1998), pp. 1–7.
- [3] ARENAS, M., BERTOSSI, L. E., AND CHOMICKI, J. Consistent query answers in inconsistent databases. In *PODS* (1999), pp. 68–79.
- [4] BELNAP, N. D. *Modern Uses of Multiple-Valued Logic*. Springer Netherlands, 1977, ch. A Useful Four-Valued Logic, pp. 5–37.
- [5] BENFERHAT, S., DUBOIS, D., AND PRADE, H. *Proc. of ECSQARU*. Springer Berlin Heidelberg, 1995, ch. A local approach to reasoning under inconsistency in stratified knowledge bases, pp. 36–43.
- [6] BERTOSSI, L., HUNTER, A., AND SCHAUB, T. *Inconsistency Tolerance*. Springer Berlin Heidelberg, 2005, ch. Introduction to Inconsistency Tolerance, pp. 1–14.
- [7] BESNARD, P., AND HUNTER, A. A logic-based theory of deductive arguments. *Artif. Intell.* 128, 1-2 (2001), 203–235.
- [8] BESNARD, P., AND SCHAUB, T. Signed systems for paraconsistent reasoning. *Journal of Automated Reasoning* 20, 1-2 (1998), 191–213.
- [9] BEX, F., PRAKKEN, H., AND REED, C. A formal analysis of the aif in terms of the aspic framework. In *Proc. of COMMA* (2010), IOS Press, pp. 99–110.
- [10] BOHANNON, P., FAN, W., FLASTER, M., AND RASTOGI, R. A cost-based model and effective heuristic for repairing constraints by value modification. In *Proc. of ACM SIGMOD* (2005), ACM, pp. 143–154.
- [11] BREWKA, G. Preferred subtheories: an extended logical framework for default reasoning. pp. 1043–1048.
- [12] CHOMICKI, J. *Proc. of ICDT*. Springer Berlin Heidelberg, 2006, ch. Consistent Query Answering: Five Easy Pieces, pp. 1–17.
- [13] DA COSTA, N. C. A., BÉZIAU, J.-Y., AND BUENO, O. A. S. Aspects of paraconsistent logic. *Logic Journal of the IGPL* 3, 4 (1995), 597–614.
- [14] DARWICHE, A., AND PEARL, J. On the logic of iterated belief revision. *Artificial intelligence* 89 (1996), 1–29.
- [15] DUNG, P. M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.* 77, 2 (1995), 321–357.
- [16] GARCÍA, A. J., AND SIMARI, G. R. Defeasible logic programming: An argumentative approach. *TPLP* 4, 1-2 (2004), 95–138.
- [17] GÄRDENFORS, P. *Knowledge in Flux: Modeling the dynamics of epistemic states*. MIT Press, 1988.
- [18] HUNTER, A. Reasoning with contradictory information using quasi-classical logic. *J. Log. Comput.* 10, 5 (2000), 677–703.
- [19] HUNTER, A. Logical comparison of inconsistent perspectives using scoring functions. *Knowledge and Information Systems* 6, 5 (2004), 528–543.
- [20] KATSUNO, H., AND MENDELZON, A. O. On the Difference between Updating a Knowledge Base and Revising it. In *Belief Revision*. 1992, pp. 183–203.
- [21] KNIGHT, K. Measuring inconsistency. *Journal of Philosophical Logic* 31, 1 (2001), 77–98.
- [22] LEHMANN, D. J. Belief revision, revised. In *Proc. of IJCAI* (1995), pp. 1534–1540.
- [23] LOZINSKII, E. L. Information and evidence in logic systems. *J. Exp. Theor. Artif. Intell.* 6, 2 (1994), 163–193.
- [24] MANOR, N. R. A. R. On inferences from inconsistent premises. *Theory and Decision* 1, 2 (1970), 179–217.
- [25] NUTE, D. Defeasible reasoning and decision support systems. *Decision Support Systems* 4, 1 (1988), 97–110.
- [26] POLLOCK, J. L. Defeasible reasoning. *Cognitive Science* 11, 4 (1987), 481–518.
- [27] REITER, R. Towards a logical reconstruction of relational database theory. In *On Conceptual Modelling (Intervale)* (1982), pp. 191–233.
- [28] SHANNON, C. E. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev.* 5, 1 (2001), 3–55.
- [29] STAWORKO, S., CHOMICKI, J., AND MARCINKOWSKI, J. Prioritized repairing and consistent query answering in relational databases. *Annals of Mathematics and Artificial Intelligence* 64, 2 (2012), 209–246.
- [30] VREESWIJK, G. A. Abstract argumentation systems. *Artificial Intelligence* 90, 1-2 (1997), 225 – 279.
- [31] WONG, P., AND BESNARD, P. Paraconsistent reasoning as an analytic tool. *Logic Journal of the IGPL* 9, 2 (2001), 217–230.

Propuesta de una Arquitectura Argumentativa Flexible en DeLP

Federico Rosenzvaig^{1,2}

Guillermo R. Simari¹

¹Laboratorio de Investigación y Desarrollo en Inteligencia Artificial, Depto. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, Bahía Blanca, Argentina.

²Depto. de Informática, Universidad Nacional de Santiago del Estero, Santiago del Estero, Argentina
e-mail: federico.rosenzvaig@cs.uns.edu.ar - grs@cs.uns.edu.ar

Resumen

Los *Sistemas Argumentativos Basados en Reglas (SABR)* son formalismos de argumentación dónde el conocimiento de un agente incluye un conjunto de reglas de inferencia a partir de las cuales se pueden construir argumentos a favor o en contra de una afirmación. Estos sistemas son de particular interés en el área de *Inteligencia Artificial (IA)* dado que este tipo de reglas de inferencia permiten representar conocimiento de sentido común, y la construcción de argumentos puede realizarse de manera automática.

El objetivo general de esta investigación es mejorar la capacidad de representación de los *SABR* brindando la posibilidad de adaptarse a las necesidades de los diferentes dominios para los cuales son dirigidos. En particular, se pretende diseñar un *SABR* con una arquitectura lo suficientemente flexible para: crear diferentes modelos argumentativos que representen una determinada situación del mundo real, y analizar dichos modelos de acuerdo a sus características sobresalientes.

Palabras clave: Sistemas Argumentativos Basado en Reglas, Inteligencia Artificial, Arquitectura Argumentativa Flexible.

Contexto

Esta línea de investigación está inserta en el marco del desarrollo de una tesis de maestría para optar por el título de Magister en Ciencias de la Computación del Departamento de Ciencias de la Computación de la Universidad Nacional del Sur. Se llevará a cabo dentro del ámbito del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA), y está asociada a los siguientes proyectos de investigación:

- “Representación de Conocimiento y Razonamiento Argumentativo: Herramientas Inteligentes para la Web y las Bases de Datos Federadas”. 24/N030, 01/01/11 – 31/12/2014.
- “Combinación de Revisión de Creencias y Argumentación para mejorar las capacidades de Razonamiento y modelado de la Dinámica de Conocimiento en Sistemas Multiagentes, PIP-CONICET (PIP 112-201101-01000), 01/01/2012 – 31/12/2014.
- “Representación de conocimiento, y Razonamiento argumentativo: Herramientas inteligentes”, 24/N030, 01/01/2011 – 31/12/2014.

Introducción

En los últimos años, la Inteligencia Artificial (IA) a puesto especial interés en imitar el

razonamiento humano frente a situaciones problemáticas. En particular, el área de la representación del conocimiento y el razonamiento rebatible que estudia el área de la argumentación rebatible, se especializa en modelar el proceso de razonamiento humano de manera tal de establecer que conclusiones son aceptables en un contexto de desacuerdo. En términos generales, las teorías de la argumentación se ocupan de analizar las interacciones entre los argumentos que están a favor o en contra de una determinada conclusión, y formar así una base de creencias que sería utilizada para afrontar las diversas situaciones problemáticas del mundo real. Estas teorías son ampliamente utilizadas en diversos ámbitos, tales como el razonamiento legal [3, 2], los sistemas de recomendación [13, 4], los agentes autónomos y sistemas multiagente [1, 12], y muchos otros [10, 16].

Los Sistemas Argumentativos Basados en Reglas (*SABR*) son formalismos de argumentación en los cuales el conocimiento de un agente incluye un conjunto de reglas de inferencia que permiten construir argumentos a favor o en contra de una afirmación. Estos sistemas son de particular interés en el área de Inteligencia Artificial (IA) dado que este tipo de reglas de inferencia son útiles para representar conocimiento de sentido común, y la construcción de argumentos puede realizarse de manera automática. Sin embargo, los *SABR* se construyen para representar un determinado tipo de formalismo, lo cual es una desventaja, ya que si se necesita utilizar otro tipo de formalismo hay que construir otro *SABR* que implemente dicho formalismo. Por ello, el objetivo general de esta investigación es incrementar la capacidad de representación de los *SABR* brindando la posibilidad de adaptarse a las necesidades de los diferentes dominios para los cuales son dirigidos. En particular, se pretende diseñar un *SABR* con una arquitectura lo suficientemente flexible para: crear diferentes modelos argumentativos que representen una determina-

da situación del mundo real, y analizar dichos modelos de acuerdo a sus características sobresalientes. En esta ocasión, analizaremos la arquitectura del razonador DeLP y trataremos de adecuarla definiendo interfaces estándar entre sus distintos módulos, con esto se obtendría una arquitectura flexible para el razonador.

Líneas de investigación y desarrollo

Actualmente, la arquitectura de un razonador argumentativo, como el utilizado en DeLP, no permite implementar de una manera sencilla dichas extensiones de una manera automática, significando así un esfuerzo considerable al reescribir el código del razonador argumentativo o utilizando diversos razonadores para modelar una misma situación problemática del mundo real.

En esta línea de investigación se estudiará como implementar extensiones a los formalismos argumentativos sobre un razonador que poseerá una arquitectura flexible para implementar automáticamente dichas extensiones de acuerdo a las preferencias del usuario. En particular, analizaremos el razonador argumentativo implementado por DeLP para crear y analizar modelos argumentativos complejos. De esta manera, el nuevo razonador nos dará la posibilidad de representar modelos argumentativos dinámicos [8], extender los esquemas de razonamiento que el mismo pueda implementar [6], considerar las valuaciones de los argumentos que dependan de las características del dominio de aplicación [5], considerar diferentes clases de ataques entre argumentos [9], entre otros.

Argumentación

La argumentación constituye uno de los principales componentes de la inteligencia hu-

mana. La habilidad de participar en discusiones es esencial para que los humanos puedan entender nuevos problemas, para llevar a cabo razonamientos científicos, expresarse, y aclarar y defender sus opiniones. Aunque en la teoría de la argumentación se incluye el debate y la negociación, las cuales están dirigidas a alcanzar unas conclusiones de mutuo acuerdo aceptables, su principal motivación es quizás el debate social en el que la victoria sobre un oponente es el principal objetivo. Este arte y ciencia es con frecuencia el medio por el cual algunas personas protegen sus creencias o propios intereses en un diálogo racional, en simples coloquios o durante el proceso de argumentación o defensa de ideas. Los investigadores dedicados al área de la argumentación estudian las racionalizaciones post hoc mediante las cuales un individuo puede justificar decisiones que originalmente pudieron haber sido realizadas de forma irracional [17].

En la actualidad, el estudio de la argumentación ha recobrado vigencia debido a la gran influencia que los medios de comunicación tienen sobre la sociedad. Esta influencia se manifiesta en el planteamiento de estrategias argumentativas para convencer al público acerca de ciertos valores e ideas. Ejemplo de esto son los discursos argumentativos relacionados con la publicidad o el pensamiento político. Así pues, la principal motivación del estudio de la argumentación, consiste en establecer si el razonamiento planteado es verosímil, es decir, si quien es objeto de la argumentación estará dispuesto a aceptarla.

Sistemas Argumentativos Basados en Reglas

En argumentación una proposición es aceptada o no de acuerdo a un análisis de las razones de las que se dispone para creer o no en la misma, donde estas razones o justificaciones toman la forma de argumentos [7]. Además, la

manera en que estos argumentos son considerados permite la automatización de este tipo de razonamiento.

En los sistemas argumentativos basados en reglas (SABR), existe un conjunto de reglas de inferencia con las cuales, a partir de cierta información (antecedente o conjunto de premisas) se puede inferir de manera tentativa nueva información (consecuente). En este tipo de sistemas, las reglas son almacenadas en una base de conocimiento, junto a otra información en forma de hechos o presuposiciones, que representan la evidencia que el agente obtiene de su entorno. A partir de esta evidencia, el agente puede usar las reglas de inferencia para construir argumentos a favor o en contra de una afirmación. Una vez hecho esto se evalúan todos los argumentos construidos, y se determina cuáles de ellos son aceptados, buscando concluir si, a partir de la base de conocimiento del agente, está afirmación puede aceptarse o no. Estos formalismos son no-monótonos dado que la introducción de nueva información al sistema puede generar nuevos argumentos que resultan contradictorios con algunos de los ya existentes.

Arquitectura Flexible de un Razonador Argumentativo

En general, en la mayoría de los SABR, argumentos y contra-argumentos son comparados utilizando un criterio de preferencia predeterminado permitiendo decidir si un ataque tiene éxito. No obstante, las capacidades de representación y el tratamiento de los conflictos pueden ser definido de diversas maneras, estableciendo así la necesidad de contar con una arquitectura más flexible que permita representar y analizar un determinado modelo argumentativo desde diferentes puntos de vista.

En los últimos años, el campo de aplicación de la argumentación se ha expandido veloz-

mente, en gran parte debido a los avances teóricos, pero también gracias a la demostración exitosa de su uso práctico en un gran número de dominios de aplicación, tales como el razonamiento legal [15], la ingeniería del conocimiento [10], los sistemas multi-agentes [14, 1], y el e-government [12], entre muchos otros [16]. Por ello, es latente la necesidad de contar con un formalismo que se adapte a las necesidades del dominio sin tener que obligar al usuario moverse de un formalismo a otro para poder modelar y analizar dichos dominios.

Objetivos y Resultados Esperados

En el LIDIA a través de los años se han llevado a cabo diferentes proyectos sobre Sistemas de Argumentación, en particular investigaciones dedicadas a desarrollar sistemas de argumentación masiva. Varias trabajos proponiendo la creación de mecanismos que pudieran mejorar la complejidad computacional de los sistemas de argumentación basados en Defeasible Logic Programming (DeLP) [11] fueron publicados en conferencias y revistas internacionales.

Esta línea de investigación tiene por objetivo general explorar y analizar la capacidad de representación de los sistemas argumentativos basados en reglas, particularmente en DeLP, con la intención de rediseñar su arquitectura para incrementar su capacidad de representación y análisis dentro de un modelo argumentativo.

Los objetivos específicos de esta línea de investigación consisten en:

- Analizar la arquitectura actual del razonador argumentativo DeLP, para poder realizar una modularización del mismo.
- Definir interfaces estándar entre los módulos del razonador para poder ampliar sus capaci-

dades, ya sean de representación o de razonamiento.

- Diseñar e implementar los módulos que permitan representar la disponibilidad temporal de los argumentos con el objetivo de crear modelos argumentativos dinámicos.
- Diseñar e implementar los módulos que permitan representar las valuaciones que tiene asociado los argumentos, y como éstas son afectadas por las interrelaciones que existen entre los mismos.

Formación de Recursos Humanos

Actualmente el equipo de trabajo de esta línea de investigación se encuentra compuesto por un estudiante de posgrado de la Universidad Nacional del Sur y su director. Por otra parte se vincula con un grupo de trabajo sobre argumentación compuesto por doctorandos e investigadores formados.

Referencias

- [1] L. Amgoud, N. Maudet, and S. Parsons. An argumentation-based semantics for agent communication languages. In *15th. ECAI*, pages 38–42, Lyon, France, 2002.
- [2] Leila Amgoud and Henri Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 173(3):413–436, 2009.
- [3] Trevor Bench-Capon and Giovanni Sartor. A model of legal reasoning with cases incorporating theories and values. *Artificial Intelligence*, 150(1):97–143, 2003.
- [4] Cristian E Briguez, Maximiliano CD Budan, Cristhian AD Deagustini, Ana G Maguitman, Marcela Capobianco, and Gui-

- llermo R Simari. Argument-based mixed recommenders and their application to movie suggestion. *Expert Systems with Applications*, 41(14):6467–6482, 2014.
- [5] Maximiliano CD Budán, Mauro Gómez Lucero, Ignacio Viglizzo, and Guillermo R Simari. A labeled argumentation framework. *Journal of Applied Logic*, 13(4):534–553, 2015.
- [6] Paola Daniela Budan, Maximiliano Budán, and Guillermo Ricardo Simari. An approach to argumentation schemes that appeal to expert opinion. *Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial*, 16(52):52–64, 2013.
- [7] Carlos Iván Chesñevar, Ana Gabriela Maguitman, and Ronald Prescott Loui. Logical models of argument. *ACM Computing Surveys (CSUR)*, 32(4):337–383, 2000.
- [8] Maria Laura Cobo, Diego C Martínez, and Guillermo Ricardo Simari. On admissibility in timed abstract argumentation frameworks. In *ECAI*, volume 215, pages 1007–1008, 2010.
- [9] Andrea Cohen, Alejandro J Garcia, and Guillermo R Simari. Extending delp with attack and support for defeasible rules. In *Advances in Artificial Intelligence—IBERAMIA 2010*, pages 90–99. Springer, 2010.
- [10] D. Robertson D. Carbogim and J. Lee. *Argument-based applications to knowledge engineering*. The Knowledge Engineering Review 15 (2), pp. 119-149, 2000.
- [11] Alejandro Javier Garcia and Guillermo Ricardo Simari. Defeasible logic programming: An argumentative approach. *TPLP*, 4(1-2):95–138, 2004.
- [12] T. J. M. Bench-Capon K. Atkinson and P. McBurney. *Multi-agent argumentation for e-democracy*. Proceedings of the Third European Workshop on Multi-Agent Systems, pp. 35-46, Brussels, Belgium, Koninklijke Vlaamse Academie, 2005.
- [13] Ana Gabriela Maguitman, Ricardo Simari Guillermo, et al. Recommender system technologies based on argumentation 1. In *Proceedings of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, pages 50–73. IOS Press, 2007.
- [14] S. Parsons, C. Sierrra, and N. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8:261–292, 1998.
- [15] H. Prakken and G. Sartor. The role of logic in computational models of legal argument - a critical survey, in. In A. Kakas and F. Sadri, editors, *Computational Logic: Logic Programming and Beyond*, pages 342–380. Springer, 2002.
- [16] I. Rahwan and G. Simari, editors. *Argumentation in Artificial Intelligence*. Springer-Verlag, 2009.
- [17] Charles Arthur Willard. *A theory of argumentation*. University of Alabama Press, 2003.

Arquitectura General de un Sistema Estructurador de Argumentos

Paola D. Budán^{1,2}

Maria Vanina Martínez^{1,3,4}

Guillermo R. Simari^{1,3}

¹Laboratorio de Investigación y Desarrollo en Inteligencia Artificial, Depto. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, Bahía Blanca, Argentina

²Depto. de Informática, Universidad Nacional de Santiago del Estero, Santiago del Estero, Argentina

³Depto. de Cs. e Ing. de la Computación, Universidad Nacional del Sur, Bahía Blanca, Argentina

⁴Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina

e-mail: pbudan@unse.edu.ar - mvm@cs.uns.edu.ar - grs@cs.uns.edu.ar

Resumen

Los esquemas de argumentación son dispositivos de razonamiento semi-estructurados empleados en conversaciones cotidianas, y en textos o discursos escritos, que permiten expresar una línea argumental. Sin embargo, dado un discurso o un texto, el proceso de encontrar el esquema de argumentación subyacente es una tarea compleja. Esta tarea requiere un sistema general que trate con diversas actividades relacionadas con el procesamiento del lenguaje natural y la minería de argumentos entre otras, y que luego sea capaz de transformar el producto de estas actividades en insumos que permitan razonar en base a la información potencialmente contradictoria existente en el discurso o texto bajo análisis. Es por este que propondremos una arquitectura general denominada *sistema estructurador de argumentos* a la que podría ajustarse cualquier desarrollo de sistemas cuyo objetivo sea añadir formalización o estructura a dichos esquemas los cuales, al vincularse con el lenguaje natural, pueden presentar un esquema de razonamiento incompleto, o tratar con información incoherente. Esta arquitectura persigue el objetivo de emular

en forma automática o semi-automática el proceso de toma de decisiones con información contradictoria, inconsistente y/o incompleta.

Palabras clave: Argumentación - Esquemas de Argumentación - Sistema Estructurador de Argumentos - Lenguaje Natural

Contexto

Esta línea de investigación está inserta en el marco del desarrollo de una tesis de doctorado para optar por el título de Doctor en Ciencias de la Computación del Departamento de Ciencias de la Computación de la Universidad Nacional del Sur, trabajo éste que se denomina “Formalización de las estructuras de los Esquemas de Argumentación”. Se llevará a cabo dentro del ámbito del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA), y está asociada a los siguientes proyectos de investigación:

- “Combinación de Revisión de Creencias y Argumentación para mejorar las capacidades de Razonamiento y modelado de la Dinámica de Conocimiento en Sistemas Multiagentes”, PIP-CONICET (PIP 112-201101-01000), 01/01/2012 – 31/12/2014.

- “Representación de conocimiento, y Razonamiento argumentativo: Herramientas inteligentes”, 24/N030, 01/01/2011 – 31/12/2014.
- “Representación de Conocimiento y Razonamiento Argumentativo: Herramientas Inteligentes para la Web y las Bases de Datos Federadas”. 24/N030, 01/01/11 – 31/12/2014.

Introducción

Los esquemas de argumentación [9, 10, 11, 12] se encuentran implícitos en los discursos en lenguaje natural por ser patrones de razonamiento semi-estructurados que permiten expresar una línea de pensamiento. Los esfuerzos para agregar formalización a los esquemas de argumentación que existen en un discurso o texto dado, pueden enfocarse en dos procesos principales:

- Determinar el patrón de razonamiento o esquema de argumentación que subyace en un discurso o texto dado, con la mayor fidelidad a la intención del proponente.
- Transformar los argumentos del discurso cotidiano expresados de manera semi-formal mediante los esquemas de argumentación, a una estructura más formal que permita manipular la información contradictoria que pueda existir en el mundo considerado o dominio del discurso.

El primer proceso mencionado está siendo abordado por diferentes líneas de investigación dentro de la Inteligencia Artificial (IA) que pueden realizar aportes significativos, puesto que se trata de una tarea altamente vinculada al procesamiento del lenguaje natural. Determinar en forma automática o semi-automática la intencionalidad de un discurso permite introducir mejoras en el proceso de inferir el(los) patrón(es) de razonamiento subyacente(s), sin embargo, es una tarea compleja de implementar. En lo que se refiere al segundo proceso,

existen herramientas desarrolladas en el área de la IA que agregan utilidad al razonamiento automático basado en argumentos, como por ejemplo DeLP[6], que permite manipular una base con conocimiento contradictorio, es decir, razones a favor y en contra de una determinada conclusión.

Abarcando estos dos procesos divididos en módulos, se propone la arquitectura general o de alto nivel de un *sistema estructurador de argumentos*, la cual será detallada en el resto del documento.

Líneas de investigación y desarrollo

La presente línea de investigación estudiará un mecanismo que permita transformar los esquemas de argumentación subyacentes en un discurso o texto dado, a una forma adecuada para razonar en base al conocimiento que de ellos se deriva, considerando que este conocimiento puede ser contradictorio. Se considera que este mecanismo se puede integrar en un sistema aún mayor que entremezcla desarrollos inherentes al lenguaje natural y a la minería de argumentos.

Arquitectura General

La arquitectura general del *sistema estructurador de argumentos* capaz de llevar a cabo los procesos anteriormente citados se muestra en la Figura 1. Los módulos representados como cajas negras aún no han sido desarrollados, y pueden ser explotados en tareas según la línea de investigación que se siga. Es decir, ante una arquitectura tan general como la propuesta, la implementación de los módulos pueden realizarse de diversas maneras.

Esta arquitectura de alto nivel se ha diseñado a los efectos de no acotar las posibilidades de implementación futura de cada módu-

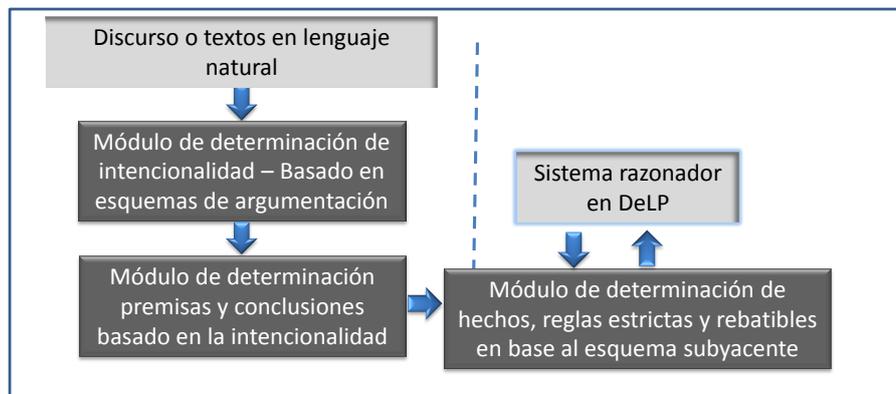


Figura 1: *Modelo de un sistema para estructurar esquemas de argumentación*

lo, ya que cada componente de la arquitectura se puede trabajar en forma totalmente independiente y utilizando avances computacionales que se pudieran producir en otras áreas de investigación.

Módulo de determinación de intencionalidad

En este módulo cobran importancia los esquemas de argumentación como formas abstractas de razonamiento comúnmente usadas en la conversación cotidiana y en los discursos escritos o textos de diferentes áreas. Constituyen una manera de argumentar, y capturan los patrones estereotipados del razonamiento humano [10, 9, 11]. La mayoría de estos esquemas son formas plausibles de razonamiento que no se ajustan a las formas tradicionales de argumentos deductivos o inductivos. Dentro de los esquemas de argumentación se encuentran: los que recurren a la opinión del experto, los que se basan en la opinión del experto, los que argumentan desde la analogía, los que tienen en cuenta la correlación entre causas-efectos, y los basados en signos, entre otros. Cada esquema está compuesto por premisas y conclusiones que configuran un patrón de razonamiento con intencionalidad explícita, y contienen además

un conjunto de preguntas claves o críticas que hacen coincidir el esquema que representan con una manera estándar de sondear críticamente un argumento para encontrar sus potenciales puntos débiles. El rol argumentativo de las preguntas críticas en estos esquemas aún no ha sido delimitado en forma precisa y clara, pero se puede decir que se han originado para evaluar la relevancia de un argumento dentro de dicho esquema [9].

El módulo de determinación de la intencionalidad de un discurso puede hacer uso de los esquemas de argumentación ya que, una vez encontrado el patrón de razonamiento subyacente en un discurso dado, se podría etiquetar dicho discurso con alguna (s) de las siguientes etiquetas: *proponente_experto*; *proponente_posicion_conocer*; *proponente_testimonial*; *opinión_popular*; *signo*; *analogía*, o la(s) que corresponda(n) según el esquema que sigue el discurso. Esto implica realizar algún tipo de minería sobre el texto con el objetivo de encontrar palabras claves inherentes a cada esquema. Independientemente de la forma en la que esta búsqueda se implemente, la concordancia entre las palabras claves y las palabras explícitas en el documento puede ser un punto de partida para resolver es-

te núcleo problemático. Sin embargo, esta solución sólo es factible cuando las palabras claves se mencionan exactamente como su patrón lingüístico. Cuando no es posible encontrar estas palabras, necesitamos explorar los indicios que nos ayuden a establecer el patrón de razonamiento subyacente en el discurso bajo análisis.

Módulo de determinación de premisas y conclusiones

Este módulo podría basarse en técnicas de minería sobre argumentos, y junto con la información que le proporcionan las etiquetas asignadas por el módulo anterior, determinar el ajuste del texto o discurso al patrón de razonamiento subyacente. Una vez determinada la intencionalidad del discurso, comienza el proceso para desagregar este discurso en argumentos que siguen un determinado patrón. Esto implica, a alto nivel, encontrar premisas y conclusiones, y luego evaluar el grado en el cual éstas se adecuan al esquema de argumentación subyacente al discurso. Entre las propuestas de investigaciones recientes que trabajan sobre la detección automática de premisas y conclusiones podemos mencionar a Cabrio[2, 3] quien propone la identificación de afirmaciones de ataque o de soporte en un diálogo, que son encontradas con la técnica de la aproximación de vinculación textual propuesta por Dagan[4]. Dicha técnica permite generar argumentos abstractos desde un texto en lenguaje natural y luego mapearlos a un módulo de argumentación abstracta como el framework de Dung[5]. De esta forma, dos argumentos pueden estar relacionados por una contradicción o por una vinculación. La contradicción se mapea con la relación de ataque y la vinculación con una relación de soporte modelada mediante ataques mediados [1].

Módulo de determinación de hechos, reglas estrictas y rebatibles

Una vez determinadas las premisas, las conclusiones y el esquema de argumentación de un texto o discurso en lenguaje natural, es deseable hacer que estos elementos adquieran utilidad práctica en el campo de la IA. Esta utilidad práctica puede estar dada por un programa que razone en base a información contradictoria, ya que en un texto o discurso en lenguaje natural encontramos argumentos y contraargumentos para un tema en particular. Entre los sistemas razonadores que trabajan con bases de conocimiento combinando la programación en lógica con la argumentación rebatible, existe la *Programación en Lógica Rebatible (DeLP, por sus siglas en inglés)*[6]. De la programación en lógica DeLP hace uso de la forma de representar el conocimiento a través de reglas, mientras que la argumentación rebatible provee los mecanismos de inferencia para decidir entre metas contradictorias. La naturaleza de estos patrones de razonamiento es muy variada dificultando la formalización de un esquema de traducción general y único que tome como entrada la descripción de una situación del mundo real, identifique la intención argumentativa o el patrón de razonamiento que gobierna dicha descripción y produzca como salida un programa DeLP, lo cual es el fin último del *Sistema Estructurador de Argumentos* que estamos diseñando. El desarrollo de este módulo se sostiene en la premisa de que la estructura de cada esquema de argumentación nos brinda herramientas para poder definir un conjunto inicial de *restricciones de contexto* como un caso particular de hechos y reglas estrictas, y también nos permite identificar otro conjunto inicial de reglas que definen el *comportamiento esquemático* rebatible de cada patrón. La decisión de considerar las *restricciones de contexto* como conocimiento seguro se fundamenta en la necesidad de enmarcar el patrón de razona-

miento para asegurar su coherencia semántica.

Resultados y Objetivos

Las investigaciones realizadas en el área de la IA inherentes al procesamiento del lenguaje natural y a la minería de argumentos están en constante desarrollo. Nuestro objetivo principal es el abordaje del módulo de determinación de hechos, reglas estrictas y rebatibles, mediante el desarrollo de las siguientes actividades algunas de las cuales están siendo ya profundizadas:

1. Elaborar una lista de palabras claves o principios que representen el comportamiento de los razonamientos.
2. Definir cada esquema de argumentación como un conjunto de reglas estrictas y rebatibles.
3. Realizar pruebas prácticas y refinamientos a partir de los resultados obtenidos con el módulo de determinación de hechos, reglas estrictas y rebatibles. Esta actividad implica evaluar el comportamiento de las traducciones de los esquemas de argumentación a reglas, tanto en forma individual como cuando los esquemas de argumentación co-existen en un discurso.

Formación de Recursos Humanos

Actualmente el equipo de trabajo de esta línea de investigación se encuentra compuesto una estudiante de posgrado de la Universidad Nacional de Bahía Blanca y su director. Por otra parte se vincula con un grupo de trabajo sobre argumentación compuesto por doctorandos e investigadores formados.

Referencias

- [1] Guido Boella, Dov M Gabbay, Leendert WN van der Torre, and Serena Villata. Support in

abstract argumentation. In *COMMA*, pages 111–122, 2010.

- [2] Elena Cabrio and Serena Villata. Generating abstract arguments: A natural language approach. In *COMMA*, pages 454–461, 2012.
- [3] Elena Cabrio and Serena Villata. Natural language arguments: A combined approach. In *ECAI*, volume 242, pages 205–210, 2012.
- [4] Ido Dagan, Bill Dolan, Bernardo Magnini, and Dan Roth. Recognizing textual entailment: Rational, evaluation and approaches—erratum. *Natural Language Engineering*, 16(01):105–105, 2010.
- [5] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*, 77(2):321–357, 1995.
- [6] A.J. García and G.R. Simari. Defeasible logic programming: An argumentative approach. *Theory and practice of logic programming*, 4(1+ 2):95–138, 2004.
- [7] Chris Reed and Glenn Rowe. Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(04):961–979, 2004.
- [8] Glenn Rowe, Fabrizio Macagno, Chris Reed, and Douglas Walton. Araucaria as a tool for diagramming arguments in teaching and studying philosophy. *Teaching Philosophy*, 29(2):111–124, 2006.
- [9] D. Walton. Justification of argumentation schemes. *Australasian journal of logic*, 3:1–13, 2005.
- [10] D. Walton. *Fundamentals of critical argumentation*. Cambridge Univ Press, 2006.
- [11] D. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, Cambridge, UK, 2008.
- [12] Douglas Walton and Chris Reed. Diagramming, argumentation schemes and critical questions. In *Anyone Who Has a View*, pages 195–211. Springer, 2003.

Entornos Argumentativos Dinámicos

Maria Laura Cobo¹ Maximiliano C. D. Budán^{1,2,3} Sergio A. Gómez¹
Diego C. Martínez^{1,2} Guillermo R. Simari¹

¹ Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA), Depto. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, Bahía Blanca, Argentina.

² Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina.

³ Depto. de Matemática, Universidad Nacional de Santiago del Estero, Santiago del Estero, Argentina.
e-mail: {mlc,mcdb,sag,dcm,grs}@cs.uns.edu.ar

Resumen

El objetivo general de esta línea de investigación es aumentar la capacidad de representación de los sistemas argumentativos (AS), estudiando y modelando el fortalecimiento y debilitamiento de los argumentos dentro de entornos dinámicos. Para ello, añadiremos un meta-nivel de información a los argumentos en la forma de etiquetas extendiendo así las capacidades de representación de la estructura. Las etiquetas serán utilizadas para introducir la representación de características especiales asociadas a los argumentos y cómo estos varían en el tiempo, la cual será utilizada en el proceso de aceptabilidad de los argumentos produciendo de esta forma resultados más refinados.

Palabras clave: Sistemas Argumentativos, Características de los Argumentos, Disponibilidad Temporal, Entornos Dinámicos.

Contexto

Esta línea de investigación se llevará a cabo dentro del ámbito del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA), y está asociada a los siguientes pro-

yectos de investigación:

- “Representación de Conocimiento y Razonamiento Argumentativo: Herramientas Inteligentes para la Web y las Bases de Datos Federadas”. 24/N030, 01/01/11 – 31/12/2014.
- “Agentes Inteligentes y Creíbles en Ambientes Interactivos Digitales”, 24/ZN22, 01/01/2011 – 31/12/2014.
- “Combinación de Revisión de Creencias y Argumentación para mejorar las capacidades de Razonamiento y modelado de la Dinámica de Conocimiento en Sistemas Multiagentes, PIP-CONICET (PIP 112-201101-01000), 01/01/2012 – 31/12/2014.
- “Representación de conocimiento, y Razonamiento argumentativo: Herramientas inteligentes”, 24/N030, 01/01/2011 – 31/12/2014.

Introducción

En los últimos años, la investigación en Inteligencia Artificial (IA) ha puesto especial interés en imitar el razonamiento humano frente a situaciones problemáticas. La argumentación constituye uno de los principales herramientas de la inteligencia humana. La habilidad de participar en discusiones es esencial para que

los humanos puedan entender y resolver nuevos problemas, para llevar a cabo razonamientos científicos, expresarse, y aclarar y defender sus opiniones. Aunque en la teoría de la argumentación se incluyen el debate y la negociación, los cuales están dirigidas a alcanzar unas conclusiones de mutuo acuerdo aceptables, su principal motivación es quizás el debate social en el que defender una posición de los ataques de un oponente es el principal objetivo. Este arte y ciencia es con frecuencia el medio por el que algunas personas protegen sus creencias o propios intereses en un diálogo racional, en simples coloquios o durante la defensa de ideas. En particular, el área de la representación del conocimiento y el razonamiento rebatible que estudia el área de la argumentación rebatible, se especializa en modelar el proceso de razonamiento humano de manera tal de establecer qué conclusiones son aceptables en un contexto de desacuerdo. En términos generales, las teorías de la argumentación se ocupan de analizar las interacciones entre los argumentos que están a favor o en contra de una determinada conclusión, y formar así una base de creencias que será utilizada para afrontar las situaciones problemáticas del mundo real.

En la actualidad, el estudio de la argumentación ha recobrado vigencia debido a la gran influencia que los medios de comunicación tienen sobre la sociedad. Esta influencia se manifiesta en el planteamiento de estrategias argumentativas para convencer al público acerca de ciertos valores e ideas. Ejemplo de esto son los discursos argumentativos relacionados con la publicidad o el pensamiento político. Así pues, la principal motivación del estudio de la argumentación, consiste en establecer si el razonamiento planteado es verosímil, es decir, si quien es objeto de la argumentación estará dispuesto a aceptarla. Por ello, es importante la capacidad de representación de los formalismos argumentativos con el objetivo de capturar de manera fiel y completa el

dominio de la argumentación. Así, en ciertas aplicaciones de la argumentación, sería interesante añadir un meta-nivel de información a los argumentos en la forma de etiquetas para extender así las capacidades de representación de las estructuras argumentales. La razón de esta extensión es que las propiedades relacionadas con la solidez lógica de un argumento no siempre son las únicas necesarias para determinar su aceptabilidad otras características dependientes del dominio de aplicación pueden ser consideradas. En los últimos años en el Laboratorio de Investigación y Desarrollo en Inteligencia Artificial de la Universidad Nacional del Sur se desarrollaron formalismos que permiten considerar meta-información dependiente del dominio de aplicación dentro del proceso de razonamiento [1, 2]. Esta meta-información está asociada a los argumentos tomando la forma de etiquetas, mejorando así su capacidad de representación. Dichas etiquetas pueden ser afectadas por las relaciones existentes entre los argumentos del modelo que describe una determinada situación problemática del mundo real. Por esta razón, se definió una estructura algebraica, llamada álgebra de etiquetas argumentales, que permite la combinación y propagación de la información asociada a los argumentos en el dominio de la argumentación. La información asociada a los argumentos a través de las etiquetas son de utilidad para diversos fines, tales como: calificar cuantitativamente y cualitativamente a los argumentos que participan de una discusión argumentativa; resolver los conflictos producidos entre los argumentos del modelo que describe la inconsistencia del mundo real a través de una noción de debilitamiento; establecer diferentes grados de aceptabilidad en base a la calidad de los argumentos que forman parte del modelo argumentativo y brindar una mejor calidad de la respuesta proporcionando información adicional de los argumentos que soportan una determinada conclusión. Asimismo, tomando como motivación los

trabajos propuestos por Cobo *et al.* [5, 4] y Godo *et al.* [7], donde se contempla la disponibilidad temporal de los argumentos, recientemente se comenzó a estudiar la formalización de una álgebra de etiquetas temporales. En particular, este estudio tiene por objetivo representar modelos argumentativos dinámicos los cuales plasman las variaciones de las cualidades asociadas a los argumentos en el tiempo, combinando etiquetas temporales con etiquetas calificativas.

Líneas de Investigación y Desarrollo

El objetivo general de esta investigación es el estudio y desarrollo de capacidades para modelar discusiones argumentativas en entornos dinámicos, de manera tal de que sea posible representar las variaciones en el tiempo de las características especiales de los argumentos participantes. Este estudio involucrará sistemas de argumentación abstracta y estructurados, como así también sus aplicaciones en el razonamiento con ontologías potencialmente inconsistentes. Los estudios se centran principalmente en los sistemas etiquetados que poseen variaciones temporales.

Argumentación

La argumentación constituye uno de los principales herramientas de la inteligencia humana. La habilidad de participar en discusiones es esencial para que los humanos puedan entender nuevos problemas, para llevar a cabo razonamientos científicos, expresarse, y aclarar y defender sus opiniones. Aunque en la teoría de la argumentación se incluyen el debate y la negociación, los cuales están dirigidas a alcanzar unas conclusiones de mutuo acuerdo aceptables, su principal motivación es quizá el debate social en el que la victoria sobre un oponente es el principal objetivo.

Este arte y ciencia es con frecuencia el medio por el cual algunas personas protegen sus creencias o propios intereses en un diálogo racional, en simples coloquios o durante el proceso de argumentación o defensa de ideas. Los estudiantes de argumentación investigan las racionalizaciones post hoc mediante las cuales un individuo puede justificar decisiones que originalmente pudieron haber sido realizadas de forma irracional [8].

Sistemas Argumentativos

En argumentación, una proposición es aceptada o no de acuerdo a un análisis de las razones de las que se dispone para creer o no en la misma, donde estas razones o justificaciones toman la forma de argumentos [3]. Además, la manera en que estos argumentos son considerados permite la automatización de este tipo de razonamiento. En los sistemas argumentativos basados en reglas (SABR), existe un conjunto de reglas de inferencia con las cuales, a partir de cierta información (antecedente o conjunto de premisas) se puede inferir de manera tentativa nueva información (consecuente). En este tipo de sistemas, las reglas son almacenadas en una base de conocimiento, junto a otra información en forma de hechos o presuposiciones, que representan la evidencia que el agente obtiene de su entorno. A partir de esta evidencia, el agente puede usar un conjunto de reglas de inferencia para construir argumentos a favor o en contra de una afirmación. Una vez hecho esto, se evalúan todos los argumentos construidos y se determina cuáles de ellos son aceptados analizando las relaciones existentes entre los argumentos. Finalmente, a partir de la base de conocimiento del agente se determina si la afirmación puede aceptarse o no. Estos formalismos son no-monótonos dado que la introducción de nueva información al

sistema puede generar nuevos argumentos que resultan contradictorios con algunos de los ya existentes. En general, en la mayoría de estos formalismos, argumentos y contra-argumentos son comparados utilizando un criterio de preferencia pre-determinado permitiendo decidir si un ataque tiene éxito.

Argumentos y Sistemas Argumentativos Etiquetados

En ciertas aplicaciones de la argumentación, será beneficioso añadir meta-nivel de información a los argumentos. Por esta razón, las etiquetas extenderán las capacidades de representación de la estructura, pudiendo de esta manera modelar los atributos asociados a los argumentos. La información brindada por las etiquetas ayudarán a refinar el proceso de aceptabilidad de los argumentos, brindando mayor información sobre el conjunto de argumentos aceptados. Para la manipulación de las etiquetas, definiremos una estructura algebraica en la cual estableceremos una colección de operadores que se utilizarán para combinar y propagar las etiquetas de acuerdo a las diferentes interrelaciones que afectan a los argumentos, como ser soporte, conflicto y agregación. Como explicamos con anterioridad, en los SABR el agente puede usar un conjunto de conocimiento y reglas de inferencia para construir argumentos a favor o en contra de una afirmación. Como resultados de esta investigación, se incorporará en los SABR el tratamiento de características especiales asociadas a la base de conocimiento que el agente posee con distintos fines, como ser, determinar la fuerza o el grado de confiabilidad de un argumento. Una vez hecho esto, se evalúan todos los argumentos construidos y se determina cuáles de ellos son aceptados analizando las relaciones existentes entre los mismos, provocando fortalecimiento o debilitamiento entre argumentos. Finalmente,

se determina si la afirmación puede aceptarse o no, y con que grado de fuerza o confiabilidad es aceptada.

Objetivos y Resultados Esperados

En el LIDIA a través de los años se han llevado a cabo diferentes proyectos sobre Sistemas de Argumentación, en particular investigaciones dedicadas a desarrollar sistemas de argumentación masiva. Varios trabajos proponiendo la creación de mecanismos que pudieran mejorar la complejidad computacional de los sistemas de argumentación basados en Defeasible Logic Programming (DeLP) [6] fueron publicados en conferencias y revistas internacionales.

El objetivo general de esta investigación es el estudio y desarrollo de capacidades para modelar discusiones argumentativas en entornos argumentativos dinámicos, de manera tal de que sea posible representar las variaciones en el tiempo de las características especiales de los argumentos participantes. Este estudio involucrará sistemas de argumentación abstracta y sistemas argumentativos estructurados, como así también sus aplicaciones en el razonamiento con ontologías potencialmente inconsistentes. Los estudios se centran principalmente en los sistemas etiquetados y los sistemas temporales. Los objetivos específicos de esta línea de investigación consisten en:

- Estudiar la formalización de sistemas argumentativos etiquetados temporales, como unión de los formalismos descritos anteriormente. Esto potencia la capacidad de modelado en razonamiento temporal.
- Introducir nuevas interpretaciones para las relaciones establecidas entre los argumentos, tales como agregación, soporte y conflicto. Estas interpretaciones se incorporarán en los desarrollos de diferentes formalismos argumentati-

vos etiquetados y en los formalismos de argumentación temporal, mejorando sus capacidades de representación.

- Modelar los efectos producidos por la agregación, el soporte y el conflicto entre argumentos etiquetados, donde la calidad de los argumentos varía con el tiempo. Destacando que dichas variaciones pueden estar sujetas a los diferentes eventos que suceden en el dominio de la aplicación.
- Implementar un sistema argumentativo valorado estructurado extendiendo las capacidades de representación de DeLP para obtener resultados experimentales con el objetivo de estudiar la eficiencia y eficacia del método argumentativo valorado.
- Buscar emplear los sistemas estructurados argumentativos etiquetados como herramientas para la implementación de un sistema de recomendación y un sistema de soporte a la toma de decisiones en donde el accionar de un agente se llevara a cabo computando la calidad de los argumentos involucrados.
- Estudiar la utilización de sistemas argumentativos temporales en el razonamiento ontológico en la web.

Los avances producidos en el marco de esta investigación serán publicados en congresos nacionales, revistas iberoamericanas y congresos internacionales.

Formación de Recursos Humanos

Actualmente el equipo de trabajo de esta línea de investigación se encuentra compuesto por investigadores, becarios de posgrado, y doctorados especializados. Por otra parte se vincula con un grupo de trabajo sobre revisión de creencias, agentes inteligentes, y sistemas de recomendaciones compuesto por doctorandos e investigadores formados.

Referencias

- [1] Maximiliano CD Budán, Mauro Gómez Lucero, Carlos Chesñevar, and Guillermo R Simari. Modeling time and valuation in structured argumentation frameworks. *Information Sciences*, 290:22–44, 2015.
- [2] Maximiliano CD Budán, Mauro Gómez Lucero, Ignacio Viglizzo, and Guillermo R Simari. A labeled argumentation framework. *Journal of Applied Logic*, 13(4):534–553, 2015.
- [3] Carlos I. Chesñevar, Ana Maguitman, and Ronald Loui. *Logical models of argument*. ACM Computing Surveys, 32(4),337-383, 2000.
- [4] Maria Laura Cobo, Diego C Martínez, and Guillermo R Simari. An approach to timed abstract argumentation. In *Proc. of Int. Workshop of Non-monotonic Reasoning*, 2010.
- [5] Maria Laura Cobo, Diego C Martínez, and Guillermo Ricardo Simari. On admissibility in timed abstract argumentation frameworks. In *ECAI*, volume 215, pages 1007–1008, 2010.
- [6] Alejandro J. García and Guillermo R. Simari. *Defeasible logic programming: An argumentative approach*, *Journal of Theory and Practice of Logic Programming*. 4 (1), pp. 95-138., 2004.
- [7] Pere Pardo and Lluís Godo. t-delp: an argumentation-based temporal defeasible logic programming framework. *Annals of Mathematics and Artificial Intelligence*, 69(1):3–35, 2013.
- [8] Charles Arthur Willard. *A theory of argumentation*. University of Alabama Press, c1989, 1989.

Herramientas para representación y razonamiento con preferencias bajo incertidumbre

Fabio R. Gallo Marcelo A. Falappa Gerardo I. Simari

Inst. de Cs. e Ing. de la Computación (Universidad Nacional del Sur-CONICET)
Av. Alem 1253, (8000) Bahía Blanca, Argentina
{fabio.gallo,mfalappa,gis}@cs.uns.edu.ar

Resumen

El continuo crecimiento en variedad y cantidad de actividades *online* que las personas realizan a través de Internet ha puesto en evidencia que existe una gran cantidad de potencial conocimiento disponible para quien esté dispuesto a analizarlo y tenga las herramientas apropiadas. Consideramos que este potencial no está siendo totalmente aprovechado debido a que las herramientas actuales de representación de conocimiento y razonamiento no están acompañando adecuadamente esta evolución. En este trabajo presentamos un análisis de los desarrollos previos que deben ser adaptados y combinados para acompañar a la evolución de Web. La presente línea de investigación involucra la combinación de un conjunto de áreas que hasta ahora han evolucionado prácticamente de manera independiente: Big Data, Web Semántica, Lenguajes ontológicos extendidos y Razonamiento acerca de preferencias.

Palabras clave: Representación de conocimiento y razonamiento, Razonamiento bajo incertidumbre probabilística, Bases de datos, Lenguajes de ontología.

1. Contexto

La presente línea de investigación se encuentra inserta en el marco del proyecto PGI 24/N035 “*Argumentación y Dinámica de Creencias para mejorar las capacidades de razonamiento y representación de conocimiento de Sistemas Multi-Agente*”, financiado por la Universidad Nacional del Sur, y el proyecto PIP-CONICET 112-201101-01000 “*Combinación de Revisión de Creencias y Argumentación para mejorar las capacidades de Razonamiento y modelado de la Dinámica de Conocimiento en Sistemas Multi-agente*”, financiado por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), ambos llevado a cabo dentro del Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur.

2. Introducción

Recientemente se ha producido una verdadera explosión en la cantidad y tipo de información que está disponible a cualquier persona con una conexión a la red Internet. Esta explosión tiene sus raíces en la World Wide Web, la cual ha revolucionado las aplicaciones de Internet, permitiéndoles a los usuarios vincular los recursos unos con otros, y organizar fácilmente el material que ellos desean hacer público. La segunda revolución llegó con el advenimiento de las aplicaciones Web, en la cual los usuarios producen su propio material, tales como en blogs o foros donde los usuarios comparten información que va desde texto plano a fotografías, audio y video. Por ello los usuarios son participantes centrales en la creación de contenido (tales como redes sociales, blogs y foros), lo cual ha llevado a la adopción del nombre Web 2.0 para señalar que ha habido una evolución en el concepto originalmente propuesto.

En los últimos años la Web tuvo una vez más un paso evolutivo, a lo que muchos se refieren como la Web Social (o Web 3.0), donde los usuarios y las relaciones entre ellos son los principales actores. Otro aspecto revolucionario que aparece en la Web 3.0 es que los datos e información básica son ahora también producidas automáticamente por máquinas; ejemplos de esto son la salida de datos por los sensores alojados en la mayoría de los teléfonos inteligentes, o por los hogares inteligentes que lentamente se están volviendo más presentes. Es en estas dos últimas “evoluciones” que el tema de los grandes datos (Big Data, en inglés) toma un rol central, dado que cualquier propuesta para el procesamiento de este tipo de información deberá ser capaz de manejar el torrente de datos que generan constantemente los usuarios y las máquinas a la par.

Más allá de los nombres adoptados y la discrepancia que puede existir en su uso, es claro que los desarrollos tecnológicos en el manejo de información ha llegado a un punto tal que es necesario investigar cómo aprovechar las capacidades ofrecidas por estos desarrollos de la mejor manera posible. En particular, nuestra investigación se enfoca en el desarrollo de mecanismos de representación de conocimiento. En el entorno de la Web Semántica y sus derivados, los mecanismos básicos utilizados para este propósito en general se denominan *len-*

guajes ontológicos, los cuales suelen ser adaptaciones de la lógica de primer orden para cumplir con objetivos de expresividad y tratabilidad computacional [7, 1].

Toda la información de usuarios disponible es una buena fuente de preferencias que los mismos expresan a su manera sobre, virtualmente, cualquier tópico; por ejemplo, preferencias sobre alimentos, indumentaria, ocio, política, etc. Dado que las preferencias pueden ser vistas como un tipo de conocimiento ontológico, y que aunque se elija representar las preferencias en forma diferenciada para poder razonar con ellas, los lenguajes ontológicos extendidos deberían ser investigados y adaptados para abordar esas cuestiones. Una manera de trabajar con preferencias es establecerlas como tuplas las cuales se corresponden con átomos básicos desde el punto de vista de la lógica clásica y lenguajes ontológicos. Esta es la perspectiva mayormente adoptada en la comunidad de Bases de Datos y, en general, especifican cómo deben ordenarse los resultados de una consulta. Esta perspectiva ha sido estudiada por casi tres décadas: a partir del trabajo fundacional de [16], en [26] se presenta un resumen de los trabajos más importantes en esta línea. En forma más o menos paralela, también se ha estado investigando la intersección entre las áreas de Bases de Datos y de Representación de Conocimiento y Razonamiento, tales como en los programas lógicos de preferencia de [15], la incorporación de preferencias en formalismos como programación con la semántica de Answer Sets [6], y mecanismos de respuesta de consultas top-k en *lenguajes ontológicos* [14]. Además de los trabajos de [?, 19] que representan una aproximación inicial a problemas específicos, no se tiene conocimiento de trabajos que combinen razonamiento probabilístico con representaciones lógicas como las propuestas aquí para razonar acerca de preferencias. A su vez, es esencial llevar a cabo una evaluación empírica extensa de los resultados obtenidos.

Por último, un proceso de razonamiento que hasta ahora no ha sido aplicado en el razonamiento acerca de preferencias es el de *argumentación*. El proceso de argumentación refleja una forma de razonamiento donde tanto la conclusión como la forma de llegar a ella pueden ser cuestionadas. Este área de estudio es de especial interés en el ámbito de la Inteligencia Artificial, principalmente porque permite razonar con información incompleta e incierta, y permite manejar inconsistencias en los sistemas basados en conocimiento. Por lo tanto, este tipo de razonamiento puede ser útil para nuestros objetivos. Como punto de partida en esta investigación se considerarán tanto los sistemas de argumentación abstractos [9] como los sistemas concretos basados en reglas. Algunos ejemplos de sistemas que han sido desarrollados en las últimas décadas son: Defeasible Logic Programming (DeLP) [13], ABA [10] y ASPIC+ [21].

3. Líneas de Investigación, Desarrollo e Innovación

A continuación haremos un breve relevamiento de los ejes de investigación, enfocándonos en los trabajos que comparten en mayor medida nuestras motivaciones.

Big Data, Web Semántica y sus evoluciones

La Web Semántica [2] es un proyecto ambicioso que fue propuesto hace poco más de una década para la integración de conocimiento en la red de documentos disponibles en Internet, comúnmente llamada la *Web*. La idea principal es lograr que el contenido que existe en dichos documentos sea accesible y procesable computacionalmente para que pueda ser compartida y reusada en forma eficiente en tareas que, en general, pueden ser resueltas con mayor o menor facilidad por humanos pero no por computadoras. Por ejemplo, con un poco de trabajo un ser humano es capaz de encontrar y reservar un hotel que cuente con las prestaciones que necesita y prefiere; sin embargo, un sistema computacional se encuentra con muchos obstáculos, tales como: (1) la existencia de varios sitios Web que ofrecen el servicio de reserva de hoteles pero usan formatos de datos muy diferentes; (2) incertidumbre en los datos, tales como inconsistencia (por ejemplo, en una parte del sitio afirma que el desayuno es gratis mientras que en otra tiene un precio asociado) o incompletitud (por ejemplo, el hotel afirma disponer de conexión a Internet, pero en ninguna parte dice si es inalámbrica o por cable); y (3) el sistema automatizado no conoce las preferencias del usuario más allá de las básicas que fueron tomadas como parte de la consulta. Por lo tanto, un sistema computacional sería incapaz de decidir entre opciones que parecen ser equivalentes pero que el usuario ordenaría de una manera específica sin dudar. Por ejemplo, si el usuario quiere un hotel cerca del centro de Mar del Plata con Internet gratis, tal vez existan muchas opciones; sin embargo, dada la información de ubicación de cada uno, el usuario elegiría aquél más cercano a la playa porque esa es una de sus actividades preferidas.

Razonamiento acerca de preferencias

El estudio de la representación y el razonamiento acerca de preferencias se ha llevado a cabo en áreas tan dispares como filosofía, lógica matemática y economía, entre otras. En la tradición adoptada en filosofía, en general, las preferencias se expresan sobre “desenlaces” (comúnmente denominados *outcomes*, en Inglés), los cuales muchas veces se corresponden con asignaciones de valores de verdad a las variables de un conjunto de fórmulas en lógica de primer orden. En este marco, el trabajo de [3] apunta a cerrar la brecha entre varios formalismos propuestos en la comunidad de Inteligencia Artificial, tales como CP-Nets [5], y aquellos estudiados en Filosofía. Desde el punto de vista del razonamiento bajo incertidumbre, no existe literatura extensa que se enfoque en la

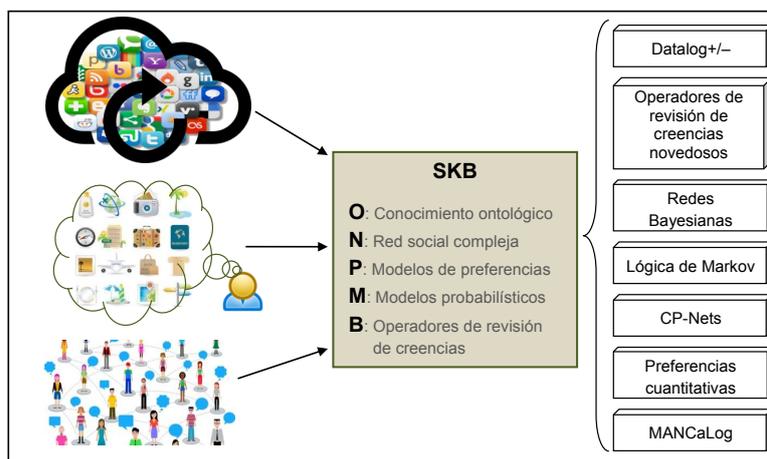


Figura 1: Las *bases de conocimiento social* (SKBs) se construyen a partir de información de la Web, como la que surge de la interacción de los usuarios con servicios *online* y redes sociales. Las componentes individuales se modelan con diferentes tipos de formalismos propuestos en la literatura para resolver problemas más específicos.

combinación de ambos; las extensiones probabilísticas de CP-Nets estudiadas en [4, 8] son dos ejemplos recientes.

Lenguajes ontológicos extendidos

Los lenguajes ontológicos, sistemas basados en reglas, y la integración de ambos es un tema central para desarrollos en la Web Semántica. Si bien existen muchas propuestas para esta integración, tales como acoplamientos débiles, fuertes e híbridos, como así también en la generalización de lenguajes ontológicos para expresar reglas, la familia de lenguajes Datalog+/- [7] es la primera propuesta que generaliza las reglas y dependencias de bases de datos para la expresión de axiomas ontológicos. El desarrollo de dicha familia, que comenzó hace pocos años, se está produciendo en un momento en el que existe gran interés en la comunidad de la Web Semántica en disponer de formalismos altamente escalables. Esta falta de desarrollo hace que exista un vacío en la disponibilidad de formalismos con la capacidad de expresar conocimiento ontológico, razonamiento bajo incertidumbre y escalabilidad similar a la lograda en los sistemas de bases de datos modernos. Las propuestas de [18, 17] dan algunos pasos en esta dirección, pero existe mucho camino que recorrer en cuanto a la comprensión de cómo lograr una integración balanceada de dichas funcionalidades; a su vez, hace falta llevar a cabo un programa de implementación y evaluación empírica riguroso, con datos tanto reales como sintéticos, para complementar las propuestas teóricas.

4. Resultados y objetivos

El objetivo es investigar los aspectos computacionales del razonamiento acerca de preferencias (de entidades que pueden ser personas o grupos enteros) bajo la existencia de incertidumbre, con un enfoque particular en la tratabilidad computacional para poder aplicar los resultados a grandes datos (Big Data, en inglés).

El objetivo particular posee dos partes: (i) El desarrollo teórico de formalismos con la capacidad de representar preferencias bajo incertidumbre, incluyendo el análisis de la complejidad computacional y los factores que influyen en ella; y (ii) El análisis empírico de dichos formalismos utilizando tanto datos reales como sintéticos. El primer objetivo permitirá la identificación de aquellos formalismos que, al menos en teoría, exhiban un buen balance entre expresividad y tratabilidad computacional. Como contrapartida, el segundo objetivo permitirá confirmar que la teoría condice con la práctica y/o identificar cuáles aspectos necesitan ser ajustados. Por lo tanto, ambos objetivos deberán ser desarrollados en forma concurrente para que los resultados de uno puedan informar a los del otro.

Como punto de partida para esta investigación, actualmente se estudian los elementos básicos que serán necesarios combinar, a saber: lenguajes ontológicos, tales como los que componen la familia de Datalog+/- [7], modelos probabilísticos como las Redes Bayesianas [20] y la Lógica de Markov [22], y métodos para el manejo de preferencias, tales como los que se analizan en [26].

Como primer abordaje a las problemáticas aquí mencionadas, recientemente elaboramos una propuesta en la cual se propone una desiderata (una lista de características deseables) para el desarrollo de lo que denominamos *Bases de Conocimiento Social* (SKB, por sus siglas en inglés) [11]. El objetivo a largo plazo de ese trabajo es derivar un framework y una metodología similar a *Ontology Based Data Access* (OBDA) [25] que esté especializado en los aspectos sociales únicos mencionados anteriormente. Esta desiderata está influenciada por propuestas anteriores en la literatura de redes complejas [23, 24], la cual establece un conjunto de criterios deseables para modelos en cascada. La figura 1 muestra un esquema de alto nivel de cómo se organiza y compone una SKB.

Esta línea de investigación está también estrechamente relacionada con otros proyectos que se están llevando a cabo en el LIDIA; sus investigadores han sido partici-

pantes centrales en diferentes proyectos de investigación sobre revisión de creencias, sistemas multi-agente, lenguajes ontológicos para razonar bajo incertidumbre, razonamiento en redes complejas, robótica cognitiva, y razonamiento acerca del comportamiento de adversarios en el marco de proyectos de modelamiento cultural computacional. Dentro de estos proyectos se desarrolló una implementación del lenguaje de argumentación estructurada *DeLP* (por sus siglas en Inglés de *Defeasible Logic Programming*) [13] y hoy cuenta con una versión disponible que permite el uso del sistema a través de la Web ¹. Se ha desarrollado también una versión llamada DeLP-Server que provee un servicio de razonamiento para sistemas multi-agentes donde diferentes agentes en host remotos pueden hacer uso de este servicio de razonamiento argumentativo [12]. Estos sistemas serán utilizados a su vez en la presente línea de investigación.

5. Formación de Recursos Humanos

Dentro de esta línea de investigación se lleva a cabo la tesis de Doctor en Ciencias de la Computación de Fabio R. Gallo, bajo la dirección de Marcelo A. Falappa y Gerardo I. Simari, en desarrollo dentro del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA) perteneciente al Instituto de Ciencias e Ingeniería de la Computación (ICIC), instituto de doble dependencia de la Universidad Nacional del Sur y CONICET. Actualmente, el LIDIA cuenta con investigadores, becarios y estudiantes de posgrado trabajando intensamente en las áreas de Razonamiento bajo Incertidumbre e Inconsistencia, Web Semántica, Razonamiento sobre Preferencias, Robótica Cognitiva, Argumentación Rebatible, Revisión de Creencias y Sistemas Multi-agente.

Referencias

- [1] BAADER, F. *The description logic handbook: Theory, implementation and applications*. Cambridge university press, 2003.
- [2] BERNERS-LEE, T., HENDLER, J., LASSILA, O., ET AL. The semantic web. *Scientific american* 284, 5 (2001), 28–37.
- [3] BIENVENU, M., LANG, J., AND WILSON, N. From preference logics to preference languages, and back. In *KR*, pp. 214–224.
- [4] BIGOT, D., FARGIER, H., MENGIN, J., AND ZANUTTINI, B. Probabilistic conditional preference networks. *arXiv preprint arXiv:1309.6817*, 72–81.
- [5] BOUTILIER, C., BRAFMAN, R. I., DOMSHLAK, C., HOOS, H. H., AND POOLE, D. Cp-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *J. Artif. Intell. Res.(JAIR)* 21 (2004), 135–191.
- [6] BREWKA, G. Preferences, contexts and answer sets. In *ICLP* (2007), p. 22.
- [7] CALÌ, A., GOTTLÖB, G., AND LUKASIEWICZ, T. A general datalog-based framework for tractable query answering over ontologies. *Web Semantics: Science, Services and Agents on the World Wide Web* 14 (2012), 57–83.
- [8] CORNELIO, C., GOLDSMITH, J., MATTEI, N., ROSSI, F., AND VENABLE, K. B. Updates and uncertainty in cp-nets. In *AI 2013: Advances in Artificial Intelligence*. Springer, 2013, pp. 301–312.
- [9] DUNG, P. M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.* 77, 2 (1995), 321–358.
- [10] DUNG, P. M., KOWALSKI, R. A., AND TONI, F. Assumption-based argumentation. In *Argumentation in artificial intelligence*. Springer, 2009, pp. 199–218.
- [11] GALLO, F. R., ABAD SANTOS, N., SIMARI, G. I., AND FALAPPA, M. A. A desiderata for modeling and reasoning with social knowledge. In *XXI Congreso Argentino de Ciencias de la Computación (Junín, 2015)* (2015).
- [12] GARCÍA, A. J., ROTSTEIN, N. D., TUCAT, M., AND SIMARI, G. R. An argumentative reasoning service for deliberative agents. In *Knowledge Science, Engineering and Management, Second International Conference, KSEM 2007, Melbourne, Australia, November 28-30, 2007, Proceedings* (2007), pp. 128–139.
- [13] GARCÍA, A. J., AND SIMARI, G. R. Defeasible logic programming: An argumentative approach. *Theory and practice of logic programming* 4, 1+ 2 (2004), 95–138.
- [14] GOTTLÖB, G., LUKASIEWICZ, T., MARTINEZ, M. V., AND SIMARI, G. I. Query answering under probabilistic uncertainty in datalog+/- ontologies. *Annals of Mathematics and Artificial Intelligence* 69, 1 (2013), 37–72.
- [15] GOVINDARAJAN, K., JAYARAMAN, B., AND MANTHA, S. Preference logic programming. In *Logic Programming, Proceedings of the Twelfth International Conference on Logic Programming, Tokyo, Japan, June 13-16, 1995* (1995), pp. 731–745.
- [16] LACROIX, M., AND LAVENCY, P. Preferences; putting more knowledge into queries. In *VLDB* (1987), vol. 87, pp. 1–4.

¹http://lidia.cs.uns.edu.ar/delp_client/

- [17] LUKASIEWICZ, T., MARTINEZ, M. V., ORSI, G., AND SIMARI, G. I. Heuristic ranking in tightly coupled probabilistic description logics. *arXiv preprint arXiv:1210.4894* (2012).
- [18] LUKASIEWICZ, T., MARTINEZ, M. V., AND SIMARI, G. I. Consistent answers in probabilistic datalog+/- ontologies. In *Web Reasoning and Rule Systems - 6th International Conference, RR 2012, Vienna, Austria, September 10-12, 2012. Proceedings* (2012), pp. 156–171.
- [19] LUKASIEWICZ, T., MARTINEZ, M. V., AND SIMARI, G. I. Probabilistic preference logic networks. In *ECAI* (2014), pp. 561–566.
- [20] PEARL, J. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1998.
- [21] PRAKKEN, H. An abstract framework for argumentation with structured arguments. *Argument and Computation* 1, 2 (2010), 93–124.
- [22] RICHARDSON, M., AND DOMINGOS, P. Markov logic networks. *Machine learning* 62, 1-2 (2006), 107–136.
- [23] SHAKARIAN, P., SIMARI, G. I., AND CALLAHAN, D. Reasoning about complex networks: A logic programming approach. *TPLP* 13, 4-5-Online-Supplement (2013).
- [24] SHAKARIAN, P., SIMARI, G. I., AND SCHROEDER, R. MANCaLog: A logic for multi-attribute network cascades. In *Proc. of AAMAS-2013* (2013).
- [25] SPANOS, D.-E., STAVROU, P., AND MITROU, N. Bringing relational databases into the semantic web: A survey. *Semantic Web* 3, 2 (2012), 169–209.
- [26] STEFANIDIS, K., KOUTRIKA, G., AND PITOURA, E. A survey on representation, composition and application of preferences in database systems. *ACM Transactions on Database Systems (TODS)* 36, 3 (2011), 1–45.

Algoritmos Genéticos Distribuidos: Heterogeneidad y Migración

German Dupuy¹, Natalia Stark¹, Fernando Sanz Troiani²,
Hugo Alfonso¹, Carlos Bermudez, Carolina Salto¹, Gabriela Minetti¹
Laboratorio de Investigación en Sistemas Inteligentes (LISI)
Facultad de Ingeniería - Universidad Nacional de La Pampa
Calle 110 Esq. 9 (6360) General Pico - La Pampa - Rep. Argentina
Te. / Fax: (02302) 422780/422372, Int. 6302
e-mail: ¹{saltoc, alfonsoh@ing.unlpam.edu.ar}

Resumen Esta línea de investigación se enfoca en las metaheurísticas distribuidas, en especial los algoritmos genéticos distribuidos, dado que permiten reducir significativamente la complejidad temporal de la resolución y mejorar la calidad de las soluciones obtenidas. Dos características sumamente importantes a la hora de resolver problemas NP-duros y NP-completos.

Por un lado, esta línea de investigación se dedica a la configuración y evaluación del desempeño de los algoritmos genéticos distribuidos ejecutados sobre plataformas heterogéneas. En este sentido surge una nueva metodología, denominada HAPA, cuyo fin es brindar una implementación eficiente y eficaz de este tipo de algoritmos.

Por otro lado, la investigación se enfoca en una nueva política migratoria de los algoritmos genéticos distribuidos, con el objetivo de mejorar su desempeño. Para ello se propone una estrategia centrada en la obligatoriedad de participación en el crossover de los individuos recibidos por medio de la migración.

Palabras claves: Metaheurísticas, algoritmos genéticos distribuidos, paralelismo, heterogeneidad, migración

Contexto

Esta línea de investigación se desarrolla en el marco de dos proyectos de investigación dirigidos por la Dra. Carolina Salto y llevados a cabo en el Laboratorio de Investigación de Sistemas Inteligentes (LISI), de la Facultad de Ingeniería

de la Universidad Nacional de La Pampa. Uno de ellos acreditado por dicha facultad y el otro por la Agencia Nacional de Promoción Científica y Tecnológica. Cabe destacar que desde hace varios años, los integrantes de estos proyectos mantienen una importante vinculación con investigadores de la Universidad Nacional de San Luis (Argentina) y de la Universidad de Málaga (España), con quienes se realizan publicaciones conjuntas.

Introducción

Los algoritmos genéticos distribuidos (dGA) son una clase de metaheurísticas, de algoritmos evolutivos en particular, cuyas soluciones se evalúan y evolucionan en paralelo. Si bien el uso de metaheurísticas logra reducir significativamente la complejidad temporal de la búsqueda, la exploración sigue consumiendo mucho tiempo a la hora de resolver problemas industriales. En consecuencia, el paralelismo es necesario no solo para reducir el tiempo de resolución, sino también para mejorar la calidad de las soluciones obtenidas [1], [2], [3], ya que cuando se usa un dGA la búsqueda puede seguir un progreso diferente al de un algoritmo genético tradicional. En tanto que, la mayoría de los resultados presentados sobre dGAs asumen que las plataformas computacionales subyacentes tienen características idénticas (ambiente homogéneo) con respecto al hardware (procesadores, memoria, red) y a los componentes de software (sistema operativo) [4], [5]. Pero la realidad es que a los laboratorios actuales les resulta muy difícil mantener hardware homogéneo y cobra importan-

cia la necesidad de configurar y evaluar el desempeño de los dGAs sobre plataformas heterogéneas.

En este sentido esta línea de investigación ataca dos frentes distintos relacionados con los dGAs. El primero estudia el desempeño y la adaptación del proceso de búsqueda de los dGAs en entornos computacionales heterogéneos. En tanto que, el segundo frente está directamente vinculado con los cambios en el proceso de búsqueda atendiendo a la topología y a la política migratoria.

Desarrollo

Actualmente, los entornos de computación distribuida se componen de muchas computadoras heterogéneas capaces de trabajar de forma cooperativa. A pesar de esto, la mayor parte del trabajo en metaheurísticas paralelas supone un hardware homogéneo como plataforma subyacente. Por un lado, nuestro trabajo se centra en la propuesta de una metodología que le permite a un algoritmo genético distribuido adaptarse para lograr una mayor eficiencia [6], usando recursos de hardware disponibles con diferente potencia de cálculo y todos ellos colaborando para resolver el mismo problema. Por otra parte, la estrategia de migración juega un rol importante en el diseño de un algoritmo genético distribuido. Motivo por el cual, también, hemos propuesto en [7] una nueva política migratoria, donde se impone que los individuos recibidos en una subpoblación por los efectos de la migración formen parte de las parejas para la recombinación.

Las investigaciones más recientes de algoritmos genéticos distribuidos sobre plataformas heterogéneas [8], [9], [10], [11], [12] se enfocan en la resolución de un determinado problema, y no en la metodología que podría aplicarse cuando es necesario usar hardware heterogéneo. En este sentido, nuestra propuesta consta de un nuevo procedimiento metodológico, y el consecuente diseño algorítmico, denominado metodología *Hardware Aware Parallel Algorithms* (HAPA) [6]. El objetivo de HAPA es facilitar una implementación eficiente y numéricamente precisa de algoritmos genéticos distribuidos sobre un conjunto de máquinas con diferentes procesadores, tamaños de memoria principal, y sistemas operativos. HAPA consta de tres

fases: (i) el cálculo de un ranking de los procesadores con el fin de conocer la plataforma (*offline*), (ii) el diseño de algoritmos derivados de la fase anterior, y (iii) la validación de un dGA (*online*). En particular se ha aplicado la metodología HAPA para regular las condiciones de parada del dGA y la frecuencia de la migración de dicho algoritmo.

En cuanto al estudio de la política migratoria de los dGA, nuestro enfoque define una estrategia relacionada con la incorporación de los individuos a una población debido al intercambio producto de la migración. La propuesta se centra en la obligatoriedad de elección de los nuevos individuos recibidos en una subpoblación como padre para la recombinación [7]. Esto marca un contraste con las reglas tradicionales de un dGA, donde el individuo recibido compite con el resto de los individuos de la subpoblación en el proceso de selección de los padres para participar en la recombinación. Esto puede ocasionar que estos individuos no se seleccionen como padres, con lo cual se estaría perdiendo el posible nuevo material genético que podrían aportar.

Resultados obtenidos

Los resultados obtenidos en los dos frentes atacados por esta línea de investigación son prometedores. Por un lado, hemos llevado a cabo una serie de pruebas con el fin de evaluar el desempeño de HAPA y compararlo con el de un dGA ejecutado en un entorno homogéneo. Al analizar la repercusión de la heterogeneidad en el rendimiento resultante de un dGA y, también, la de su eficiencia en el tiempo de ejecución se concluye que: la calidad de la solución es comparable con la obtenida usando una plataforma homogénea, pero lo más importante es que se logra una disminución del esfuerzo numérico y, consecuentemente, del tiempo de ejecución para hallar dichas soluciones cuando HAPA se aplica en ambientes heterogéneos. En resumen, este trabajo aporta una forma de evitar el diseño *ad-hoc* de algoritmos para plataformas heterogéneas, también de abre una nueva línea de investigación sobre la forma de inyectar conocimientos de hardware en los parámetros del software de los algoritmos. Además de ser una propuesta innovadora, se ha

demostrado que es numéricamente competitiva y que los tiempos de ejecución son reducidos.

Por otra parte el dGA, cuya estrategia de migración considera como padre para la recombinación a aquellos nuevos individuos ingresados por una operación de migración, ha sido evaluado en un amplio conjunto de instancias del problema de planificación de proyectos de software y comparado con un dGA tradicional. Los resultados indican que además de encontrar soluciones de buena calidad, el uso de esta estrategia permite encontrarla en una menor cantidad de evaluaciones que un dGA tradicional.

Formación de recursos humanos

Cada año se incorporan al proyecto alumnos avanzados en la carrera Ingeniería en Sistemas, quienes trabajan en temas relacionados a la resolución de problemas de optimización usando técnicas inteligentes, con el objeto de guiarlos en el desarrollo de sus tesis de grado y, también, de formar futuros investigadores científicos. Por otra parte, los docentes-investigadores que integran el proyecto realizan diversos cursos de posgrado relacionados con la temática del proyecto, con el objetivo de sumar los créditos necesarios para cursar carreras de posgrado.

REFERENCES

- [1] E. Alba and M. Tomassini, "Parallelism and evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 5, pp. 443–462, Oct 2002.
- [2] E. Cantu-Paz, *Efficient and Accurate Parallel Genetic Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers, 2000.
- [3] E. Alba, *Parallel Metaheuristics: A New Class of Algorithms*. Wiley, 2005, ISBN: 978-0-471-67806-9. [Online]. Available: <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471678066.html>
- [4] E. Alba, A. J. Nebro, and J. M. Troya, "Heterogeneous computing and parallel genetic algorithms," *Journal of Parallel and Distributed Computing*, vol. 62, no. 9, pp. 1362–1385, 2002.
- [5] G. Luque and E. Alba, *Parallel Genetic Algorithms: Theory and Real World Applications*. Springer, 2011, vol. 367.
- [6] C. Salto and E. Alba, "Adapting distributed evolutionary algorithms to heterogeneous hardware," *Trans. Computational Collective Intelligence*, vol. 19, pp. 103–125, 2015.

- [7] G. Dupuy and C. Salto, "Una estrategia para la selección de individuos recibidos en la migración en un algoritmo genético distribuido," in *Proceedings of ASAI 2015 Argentine Symposium on Artificial Intelligence, 44th Argentine Conference on Informatics*, 2015, pp. 96–104.
- [8] J. W. B. Jr. and S. V. Kumar, "Asynchronous genetic algorithms for heterogeneous networks using coarse-grained dataflow," in *GECCO*, ser. Lecture Notes in Computer Science, E. Cantu-Paz, J. A. Foster, K. Deb, L. Davis, R. Roy, U.-M. O'Reilly, H.-G. Beyer, R. K. Standish, G. Kendall, S. W. Wilson, M. Harman, J. Wegener, D. Dasgupta, M. A. Potter, A. C. Schultz, K. A. Dowland, N. Jonoska, and J. F. Miller, Eds., vol. 2723. Springer, 2003, pp. 730–741. [Online]. Available: <http://dblp.uni-trier.de/db/conf/gecco/gecco2003-1.html#BaughK03>
- [9] M. García-Arenas, J. J. Merelo Guervós, P. Castillo, J. L. J. Laredo, G. Romero, and A. M. Mora, "Using free cloud storage services for distributed evolutionary algorithms," in *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, ser. GECCO '11. New York, NY, USA: ACM, 2011, pp. 1603–1610. [Online]. Available: <http://doi.acm.org/10.1145/2001576.2001792>
- [10] Y. Gong, M. Nakamura, and S. Tamaki, "Parallel genetic algorithms on line topology of heterogeneous computing resources," in *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*, ser. GECCO '05. New York, NY, USA: ACM, 2005, pp. 1447–1454. [Online]. Available: <http://doi.acm.org/10.1145/1068009.1068239>
- [11] K. Meri, M. G. Arenas, A. M. Mora, J. J. Merelo, P. A. Castillo, P. García-Sánchez, and J. L. J. Laredo, "Cloud-based evolutionary algorithms: An algorithmic study," *Natural Computing*, vol. 12, no. 2, pp. 135–147, 2012. [Online]. Available: <http://dx.doi.org/10.1007/s11047-012-9358-1>
- [12] S. Mostaghim, J. Branke, A. Lewis, and H. Schmeck, "Parallel multi-objective optimization using master-slave model on heterogeneous resources," in *IEEE Congress on Evolutionary Computation*. IEEE, 2008, pp. 1981–1987. [Online]. Available: <http://dblp.uni-trier.de/db/conf/cec/cec2008.html#MostaghimBLS08>

Tratamiento de Metas en Planificación Continua

Andrea Vicente

Gerardo Parra

email: vicenteandrea89@gmail.com, gparra@fi.uncoma.edu.ar

Grupo de Investigación en Lenguajes e Inteligencia Artificial

Departamento de Teoría de la Computación

Facultad de Informática

UNIVERSIDAD NACIONAL DEL COMAHUE

Buenos Aires 1400 - (8300)Neuquén - Argentina

Resumen

El proyecto de investigación *Agentes inteligentes en ambientes dinámicos* tiene como objetivo general el estudio y desarrollo de técnicas de Inteligencia Artificial para dotar de inteligencia y conocimiento a agentes inmersos en mundos virtuales, interactivos y dinámicos.

Diferentes líneas de investigación confluyen en este proyecto. Entre ellas se encuentran planificación, tecnologías del lenguaje humano, ingeniería de conocimiento y juegos.

El presente trabajo se enmarca en la línea planificación. El objetivo final es desarrollar un módulo de tratamiento y gestión de metas para maximizar las características de autonomía de un agente planificador en un ambiente dinámico.

Palabras Clave: AGENTES INTELIGENTES, PLANEAMIENTO, PLANIFICACIÓN CONTINUA, TRATAMIENTO DE METAS.

Contexto

Este trabajo está parcialmente financiado por la Universidad Nacional del Comahue, en el contexto del proyecto de investigación *Agentes inteligentes en ambientes dinámicos (04/F006)*.

El proyecto de investigación tiene prevista una duración de cuatro años, desde enero del 2013 hasta diciembre de 2016.

1. Introducción

En todo ambiente relevante, el dinamismo del entorno provoca, usualmente, que los agentes deban adaptarse a nuevas situaciones y resolver problemas impredecibles para poder alcanzar los objetivos finales. Por otra parte, los agentes suelen tener restricciones de tiempo para adecuarse a los cambios constantes en el entorno en el que actúan.

La Planificación Continua[10] es considerada una alternativa para trabajar en ambientes no determinísticos o cambiantes. Bajo esta alternativa, el agente se encuentra en un entorno dinámico formulando nuevas metas, planificando y actuando de manera constante. Un planificador continuo como el propuesto en [7] brinda al agente un comportamiento tanto deliberativo como reactivo, dependiendo de la situación.

La mayor parte de los mecanismos para el tratamiento de metas en ambientes dinámicos plantean que existen restricciones y condiciones sobre las metas que el agente debe evaluar[4, 6, 9]. El agente debe tener autonomía para saber cómo actuar ante estos

eventos inesperados[14]. Por ejemplo, consideremos un agente que entrega paquetes. Si la dirección donde debe entregar el paquete no existe más, entonces el agente debería ser capaz de buscar alguna alternativa.

En [13] presentamos tres estrategias de tratamiento de metas en planificación continua, estas son: Transformación de Metas[2, 12], Generación de Metas[3] y Gestión de Metas[5, 1]. El estudio de estas estrategias de tratamiento de metas para agentes planificadores en ambientes dinámicos apunta a maximizar las características de autonomía del agente y de adaptación al ambiente en que actúa. El objetivo de este trabajo es presentar las diferencias y similitudes que encontramos entre las estrategias analizadas.

Las estrategias de gestión y generación de metas asocian a cada objetivo una medida de importancia o prioridad (la gestión de metas también asocia una medida de urgencia). La transformación de metas no contempla características adicionales para las metas, pero calcula un conjunto de metas transformadas relevantes sobre la original. Por esto último se dice que bajo este enfoque las metas son *mutables* porque pueden transformarse en otras similares dependiendo de los cambios del entorno.

La gestión de metas permite al agente seleccionar aquellas metas que le proporcionan una mayor utilidad. Esto es, las metas son analizadas por medio de los atributos asociados y de esta manera llegan al conjunto de *metas activas*. La transformación de metas recibe los objetivos, y en el caso de que los cambios en el ambiente provoquen que ya no sean alcanzables, aplica una transformación buscando siempre una utilidad equivalente.

Por otro lado, la generación de metas establece que un deseo se convierte en una meta cuando es reunido con otras condiciones que hacen posible que el agente actúe. Bajo este enfoque se distinguen dos conjuntos de proposiciones atómicas, el conjunto D de todos los posibles deseos y el conjunto K de todos los posibles ítems de conocimiento. Un deseo no es una pieza de conocimiento y viceversa. De todas formas, los deseos pueden

depender de los conocimientos, mientras que los conocimientos nunca pueden depender de los deseos. La generación de metas determina las metas a adoptar considerando tanto la confiabilidad del origen de la información, así como analizando la *relevancia* de los conocimientos. El término relevancia representa la importancia que tiene un conocimiento sobre otro para adoptar un deseo. Supongamos que queremos comprar una heladera (ch) y que dicha decisión está condicionada a que tenga freezer (tf) y sea de color gris (hg). Si tf es más relevante que hg , entonces aunque la heladera sea blanca, siempre que tenga freezer, se asumiría la meta.

La generación de metas analiza cada deseo y las condiciones actuales para determinar si puede adoptar una meta; podríamos decir que determina el conjunto de metas con el que va a trabajar. En el caso de la transformación de metas, si bien tiene el conjunto de metas a llevar adelante puede decidir el conjunto de metas transformadas. La gestión de metas trabaja con un conjunto de metas, ya sea descartando, suspendiendo o reactivando las mismas durante el proceso. Para realizar esta tarea se basa en los atributos asociados a las metas que varían de acuerdo a las condiciones actuales del ambiente y a las necesidades del agente.

Los tres enfoques buscan garantizar un buen uso de recursos dado que no se planificará para metas que se crean no realizables con las condiciones actuales del ambiente. Siempre buscan generar la mayor utilidad para el agente. A su vez no hay restricciones respecto a la relación que existe entre las metas, es decir pueden ser independientes o estar interrelacionadas.

Otro enfoque también interesante y que puede ser considerado como una combinación de transformación y de generación de metas se denomina *formulación de metas*. La formulación de metas[14, 11] establece que muchos agentes autónomos, ante cambios imprevistos en el ambiente, tratan de replantear aquellas metas que originalmente les fueron dadas. Lo que se busca es tener un agente que

sea capaz de crear sus propias metas, es decir que lleve a cabo una *formulación de metas*.

La formulación de metas proporciona al agente la capacidad de tener en cuenta necesidades internas, en lugar de solo perseguir aquellas metas que le fueron dadas al inicio (por un humano). El agente puede automáticamente generar óptimos comportamientos en situaciones que no fueron consideradas por el diseñador. Decimos que tiene características de ambos enfoques dado que el agente puede replantear metas, consiguiendo metas similares (transformación de metas) y también puede generar sus propias metas en base a sus percepciones (generación de metas).

2. Líneas de Investigación y Desarrollo

El agente que implementa el Planificador Continuo se encuentra detallado en [8]. La investigación actual trata de dotar a este agente de características que le permitan comportarse de una manera más real o conveniente ante situaciones imposibles de anticipar y que lo obligen a alterar el objetivo originalmente perseguido.

Específicamente, ya se está trabajando en integrar, al planificador existente, un módulo que implementa el mecanismo de Gestión de Metas. El mismo está siendo implementado en el lenguaje Java. El módulo tiene dos modos de funcionamiento: Conservador y Arriesgado. En el modo Conservador, para determinar si una meta es emergente o no, se utiliza principalmente la medida de urgencia asociada. Además, como plantea la bibliografía consultada, se consideran las medidas de importancia para resolver los conflictos que puedan originarse entre metas con atributos prácticamente iguales.

Por otro lado, en el modo Arriesgado se considera en primer lugar la importancia asociada a la meta para resolver si la misma resulta emergente o no.

3. Resultados Esperados

El resultado de esta línea de investigación intenta mejorar el comportamiento de un agente planificador. Un tratamiento de metas adecuado y el consecuente aumento de autonomía y adaptación, permitirá que el agente planificador pueda actuar de manera superior frente a los eventos inesperados que pueden ocurrir en el ambiente en el que se desempeña. Si bien en un principio se plantea comenzar con la implementación de un enfoque de Gestión de Metas, en el futuro se espera establecer alguna combinación con los restantes enfoques analizados.

Los enfoques planteados son parte de una tesis de grado que se está desarrollando y se encuentra en etapa de redacción. Se estima finalizar este trabajo en el transcurso del corriente año.

4. Formación de Recursos Humanos

En el transcurso de este año, se espera la finalización de dos tesis de Licenciatura dirigidas por los integrantes del proyecto. Asimismo, se prevee el inicio de dos tesis de grado dirigidas y/o co-dirigidas por los miembros del grupo.

Finalmente, se espera la consolidación como investigadores de los miembros más recientes del equipo de investigación.

Referencias

- [1] A. Sloman and D. Hogg and G. Humphreys and D. Partridge and A. Ramsay and Luc P. Beaudoin and Aaron Sloman. A study of motive processing and attention. In *Prospects for Artificial Intelligence*, pages 229–238. IOS Press, 1993.
- [2] M. T. Cox and M. M. . Veloso. Goal transformations in continuous planning.

- In *In M. desJardins (Ed.)*, pages 23–30, Menlo Park, Calif, 1998. AAAI Fall Symposium on Distributed Continual Planning.
- [3] C. da Costa Pereira and A. G. B. Tetamanzi. Goal generation with relevant and trusted beliefs. In *7th international joint conference on Autonomous agents and multiagent systems*, 2008.
- [4] R. Feldmann, G. Brewka, and S. Wenzel. Planning with Prioritized Goals. In *Tenth International Conference on Principles of Knowledge Representation and Reasoning*, 2006.
- [5] Hanheide, Marc and Hawes, Nick and Wyatt, Jeremy and Göbelbecker, Moritz and Brenner, Michael and Sjö, Kristoffer and Aydemir, Alper and Jensfelt, Patric and Zender, Hendrik and Kruijff, Geert-Jan. A Framework for Goal Generation and Management. 2010.
- [6] Minh B. Do. Planning with goal utility dependencies. In *In Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI 2007)*, pages 1872–1878, 2007.
- [7] M. Moya and C. Vaucheret. Planificador continuo como controlador de agentes robots. In *X Workshop de Investigadores en Ciencias de la Computación*, General Pico, La Pampa, Argentina, 2008. Universidad Nacional de la Pampa.
- [8] Moya Mario. Control de Agentes basado en Planificacion Continua. Tesis para la Carrera Licenciatura en Ciencias de la Computación. Universidad Nacional del Comahue.
- [9] Ronen I. Brafman and Yuri Chernyavsky. Planning with Goal Preferences and Constraints. In *AAAI*, pages 182–191, 2005.
- [10] S. Russell and P. Norvig. *Artificial Intelligence: A modern approach*. Prentice Hall, New Jersey, third edition, 2009.
- [11] Swaroop Vattam, Matthew Klenk, Matthew Molineaux, David W. Aha. Breadth of Approaches to Goal Reasoning: A Research Survey. In *Annual Conference on Advances in Cognitive Systems: Workshop on Goal Reasoning*, College Park, MD: University of Maryland, Department of Computer Science, 2013. ACS Workshop (Technical Report CS-TR-5029).
- [12] M. T. Cox and C. Zhang. Mixed Initiate Goal Manipulation. *IA Magazine*, 28(2).
- [13] A. Vicente and G. Parra. Planificación continua: Análisis de estrategias de tratamiento de metas. In *XVII Workshop de Investigadores en Ciencias de la Computación*, Salta, Salta, Argentina, 2015. Universidad Nacional de Salta.
- [14] Wilson, M., Molineaux, M., & Aha, D.W. Domain-Independent Heuristics for Goal Formulation. In *Twenty-Sixth Florida Artificial Intelligence Research Society Conference*, pages 160–165, St. Pete Beach, 2013. AAAI Press.

Servicios Web combinados con Agentes Inteligentes

Gregorio N. Tkachuk, Carlos E. Vega Ugozzoli, Diana Palliotto

Instituto de Investigación en Informática y Sistemas de Información (IISI)
Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero
Av. Belgrano (S) 1912, Santiago del Estero
Teléfono: +54 385 4509500 interno 1838
gregorio@unse.edu.ar; cvegaugozzoli@hotmail.com; dpalliot@unse.edu.ar

Resumen

Actualmente, los servicios web y los agentes inteligentes constituyen dos tecnologías fundamentales para el desarrollo de aplicaciones basadas en la web, particularmente para aquellas aplicaciones en dominios sumamente dinámicos que provocan cambios vertiginosos o imprevistos en cualquier momento y afectan al sistema. Si bien las potenciales ventajas de la interoperabilidad de ambas tecnologías se reconocen ampliamente, llevar a cabo su integración no es una tarea sencilla si se tiene en cuenta que cada una de ellas se ha desarrollado considerando diferentes estándares y especificaciones. En este trabajo se presenta una línea de investigación que propone reconocer y especificar los beneficios que se pueden obtener al combinar ambas tecnologías en el desarrollo de aplicaciones web, y determinar los contextos en los cuales es ventajosa tal integración.

Palabras clave: servicios web; agentes inteligentes; aplicaciones web.

Contexto

Este trabajo presenta una de las líneas de investigación que se desarrolla dentro del proyecto denominado "Métodos, Técnicas y Herramientas para la Ingeniería de Software Orientada a Agentes" (Código 23/C096), aprobado y financiado por el Consejo de Investigaciones Científicas y Tecnológicas de la Universidad Nacional de Santiago del Estero (CICyT, UNSE) para el período 2012-2016. El proyecto forma parte del programa "Sistemas de Información Web Basados en Tecnología de Agentes", también aprobado por CICyT, UNSE.

Introducción

El avance tecnológico en las comunicaciones y la constante evolución de las plataformas para el desarrollo de software, representan un desafío para la integración de aplicaciones ya existentes con sistemas que se desarrollan con nuevas tecnologías. Esta integración debe superar una serie de diferencias para llegar a ser efectiva, ya que las nuevas tecnologías se distinguen de las existentes en varios aspectos, tales como el lenguaje de programación, el lenguaje y los

protocolos de comunicación, la plataforma de desarrollo, etc. Para superar estas divergencias es necesario recurrir a métodos estandarizados y bien definidos que posibiliten la interacción entre las distintas aplicaciones que se desarrollaron con diferentes tecnologías.

En este contexto, dos tecnologías que han ganado una creciente importancia en los últimos años son los *servicios web* (SW) y los *agentes inteligentes* (AI). Ambas tecnologías han mostrado significativos desarrollos en la definición de estándares y protocolos, y de aplicaciones industriales, pero esos avances se produjeron mayormente en forma independiente una de la otra [1, 2, 3, 4].

A partir de los SW, los millones de personas que acceden a la web para compartir información y servicios en línea, pueden ver a Internet ya no sólo como una red de transmisión de contenidos, sino que ésta se convierte en una verdadera red de ejecución de cálculos, procesamiento de transacciones comerciales y aplicaciones de negocios. Estas tareas se tornan cada vez más habituales, lo que hace imposible que un usuario determine manualmente, en el tiempo adecuado y con eficiencia, los servicios que requiere para satisfacer una necesidad concreta.

Por otro lado, la tecnología de los AI se estudia y perfecciona desde hace varios años, con aplicaciones en distintas áreas de interés, desde el desarrollo de sistemas para edificios inteligentes, dispositivos móviles, conducción autónoma de vehículos, sistemas de energía, interfaces de usuario inteligentes, sistemas de manufacturación, control de tráfico aéreo y aplicaciones espaciales, entre otras [4, 5, 6, 7].

Sin embargo, el área donde mayores esfuerzos de investigación se están realizando es en la del desarrollo de

sistemas inteligentes para aplicaciones web. Aquí, los AI y los sistemas multiagente se plantean como herramientas fundamentales para abordar los problemas de la heterogeneidad de los datos, de la distribución y el dinamismo que presenta la web [8, 9]; es decir, obtener *servicios web inteligentes*. Es importante considerar que la principal diferencia entre un *servicio web* y un *servicio web inteligente* es que este último considera la fusión entre los SW y los AI como el medio para implantar la "inteligencia" en los SW [4].

Debido a la relevancia que los SW y los AI han ganado a lo largo del tiempo, no es de extrañar que comenzaran a surgir distintas propuestas para su integración, de manera tal de obtener las mejores características y ventajas de ambas tecnologías. Sin embargo, este no es un aspecto trivial si se considera que ambos enfoques se han desarrollado con diferentes estándares y especificaciones, como es el caso de las propuestas de FIPA [10] para los AI y de W3C [11, 12] para los SW.

Líneas de Investigación, Desarrollo e Innovación

Esta línea de investigación se propone, principalmente, realizar un aporte para identificar y mostrar los beneficios que se pueden obtener de la combinación de los SW y los AI en distintas áreas de aplicación, tales como: ambientes de aprendizaje, búsquedas en la web, seguridad informática, estudios climatológicos. Esto implica:

- Desarrollar un marco que permita evaluar la calidad de la integración de SW y AI.
- Diseñar e implementar una aplicación que integre SW y AI.

- Evaluar la integración en base al marco elaborado anteriormente.

Resultados y Objetivos

Los objetivos generales que se persiguen en esta línea de investigación son los siguientes:

- Contribuir al mejoramiento en la resolución de problemas que pertenezcan de un mismo dominio, utilizando SW y AI.
- Determinar los ámbitos en los que es conveniente y ventajoso emplear SW y AI en el desarrollo de aplicaciones.

Los resultados que se esperan obtener son:

- Diseño del prototipo de una aplicación que permita la integración entre los SW y los AI.
- Elaboración de un marco de evaluación para el prototipo.
- Evaluación del prototipo y de la calidad de la integración.

Formación de Recursos Humanos

El equipo de trabajo de este proyecto está compuesto por un docente investigador formado (categoría III) y dos docentes investigadores en formación (categorías IV e inicial).

Se prevé la incorporación al equipo de un integrante estudiante que desarrollará su trabajo final de grado para obtener el título de Licenciado en Sistemas de Información dentro de esta línea de investigación.

Referencias

- [1] Erl, T. (2005). *Service-oriented architecture: concepts, technology, and design*. Pearson Education India.

[2] Erl, T. (2004). *Service-oriented architecture: a field guide to integrating XML and web services*. Prentice Hall PTR.

[3] Gomez-Sanz, J. J., & Fuentes-Fernández, R. (2015). Understanding Agent-Oriented Software Engineering methodologies. *The Knowledge Engineering Review*, 30(04), 375-393.

[4] Marin, C. A., Monch, L., Leitao, P., Vrba, P., Kazanskaia, D., Chepegin, V., ... & Mehandjiev, N. (2013, October). A conceptual architecture based on intelligent services for manufacturing support systems. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on* (pp. 4749-4754). IEEE.

[5] Zhong, N., Liu, J., & Yao, Y. (2007). Envisioning intelligent information technologies through the prism of web intelligence. *Communications of the ACM*, 50(3), 89-94.

[6] Weerawarana, S., Curbera, F., Leymann, F., Storey, T., & Ferguson, D. F. (2005). *Web services platform architecture: SOAP, WSDL, WS-policy, WS-addressing, WS-BPEL, WS-reliable messaging and more*. Prentice Hall PTR.

[7] Vrba, P., Marik, V., Siano, P., Leitão, P., Zhabelova, G., Vyatkin, V., & Strasser, T. (2014). A review of agent and service-oriented concepts applied to intelligent energy systems. *Industrial Informatics, IEEE Transactions on*, 10(3), 1890-1903.

[8] Chase, N. (2006). *Understanding web services specifications: SOAP*. IBM Corporation.

[9] Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. *ACM Sigkdd Explorations Newsletter*, 2(1), 1-15.

- [10] Chiariglione, L. (2001). FIPA: Foundation for Intelligent Physical Agents. In *FIPA Board of Directors, Resolutions of the Osaka Meeting*. [Links].
- [11] World Wide Web Consortium. Disponible en: <http://www.w3c.org>. Acceso: 23/07/2015.
- [12] Bieberstein, N. (2006). *Service-oriented architecture compass: business value, planning, and enterprise roadmap*. FT Press.

Representación y razonamiento con bases de conocimiento probabilísticas

Natalia Abad Santos[†] Gerardo I. Simari[‡]

[†] Inst. de Cs. e Ing. de la Computación (Universidad Nacional del Sur–CONICET)
Av. Alem 1253, (8000) Bahía Blanca, Argentina
gis@cs.uns.edu.ar

[‡]Departamento de Matemática, Universidad Nacional del Sur
Av. Alem 1253, (8000) Bahía Blanca, Argentina
nasantos@uns.edu.ar

Resumen

Esta línea de investigación busca realizar un relevamiento lo más abarcativo posible de los formalismos desarrollados para la representación de conocimiento probabilístico, y sus algoritmos de razonamiento asociados. El principal desafío de este trabajo – y por ende la novedad y utilidad del mismo – yace en la necesidad de reconciliar desarrollos y estudios realizados en áreas diferentes, tales como bases de datos, representación de conocimiento y razonamiento, machine learning, y otras en las que se desarrollan sistemas aplicados específicos.

Palabras clave: Representación de conocimiento y razonamiento, Razonamiento bajo incertidumbre probabilística, Bases de datos, Lenguajes de ontología.

1. Contexto

La presente línea de investigación se encuentra inserta en el marco del proyecto PRH-2014-0007 “*Respuesta de consultas a bases de conocimiento sociales con incertidumbre utilizando las preferencias de los usuarios*”, coordinado por la Agencia Nacional de Promoción Científica y Tecnológica, Ministerio de Ciencias, Tecnología e Innovación Productiva, y llevado a cabo dentro del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur.

2. Introducción

El razonamiento bajo incertidumbre es una parte central de nuestras vidas como agentes razonadores. La incertidumbre puede surgir de varios factores: vivimos en un mundo que cambia constantemente, con información frecuentemente contradictoria o incompleta, y que también contiene elementos inherentemente inciertos como el clima, el comportamiento de los demás agentes, o de sistemas complejos como la economía. Este tipo de razonamiento es entonces central para la Inteligencia Artifi-

cial, y su desarrollo y estudio adecuado es esencial para el avance de la disciplina.

Antecedentes

La *lógica probabilística*, también conocida como *lógica de probabilidad*, tiene una larga historia que se remonta a la época de Leibniz (fines del siglo 17, principios del siglo 18). En su artículo de 1984 [13], Hailperin relata esta historia con cierto detalle, y también releva el trabajo de muchos lógicos y matemáticos involucrados en el desarrollo de lógicas de probabilidad que condujeron a los formalismos de hoy en día.

De Leibniz a Boole. Tal y como se afirmó en [13], Leibniz ya visualizaba una “nueva clase de lógica” que involucraría grados de probabilidad. Como el nacimiento de la teoría de probabilidad ocurrió durante su vida, esta propuesta no fue clara, pero fue claro que visionaba un medio para estimar la probabilidad de los valores de verdad, y una teoría de prueba que sólo llevó a la probabilidad en lugar de a la certeza. No fue hasta el siglo 19 que la lógica moderna fue desarrollada y fue Boole [3] quien combinó por primera vez la lógica de “no”, “y” y “o” con la teoría de probabilidad. Hubo algunas restricciones (quizás innecesarias) en el formalismo propuesto, ya que el conectivo “o” fue interpretado como excluyente (y entonces las probabilidades de las disyunciones eran el resultado de sumar las probabilidades de sus componentes), y todos los eventos representados por proposiciones eran supuestos independientes (y entonces las probabilidades de las conjunciones eran el producto de las probabilidades de sus componentes). Este fue ciertamente el primer progreso en lógica por probabilística. Algún tiempo después en el siglo 19, C.S. Peirce también mostró gran interés en este proyecto, pero su tratado de *inferencia probable* [30] en realidad no hace un tratamiento formal de este tipo inferencias. Las lógicas multi-valuadas fueron desarrolladas en la década de 1920, y la probabilidad fue asociada naturalmente con este nuevo formalismo. Reichenbach fue uno de los más distinguidos partidarios de esta asociación [31]; Tarski, por otro

lado, se opuso a la extensión de lógicas multivaluadas incorporando probabilidad, y también manifestó que la lógica probabilística de Reichenbach no era realmente una lógica multivaluada sino un caso especial de lógica bivaluada [39].

Avances modernos. Hailperin [13] desarrolló una lógica probabilística en el nivel proposicional que generaliza levemente la idea de tabla de verdad. La noción de probabilidad es parte de la semántica, y, de esta manera, el significado de la probabilidad juega el mismo rol que la verdad en la lógica clásica. Esta lógica no está dada sintácticamente, sino semánticamente a través de la definición de *consecuencia lógica* usando la noción de *modelo probabilístico*. En la misma dirección (y publicado con sólo dos años de diferencia), Nilsson [29] propuso una generalización semántica de la lógica en la cual los valores de verdad de las sentencias son valores probabilísticos. La generalización de Nilsson se aplica a cualquier sistema lógico para el cual la consistencia de un conjunto finito de sentencias pueda ser establecida. La semántica de esta lógica probabilística asigna distribuciones de probabilidad sobre el conjunto de mundos posibles (es decir, subconjuntos de las bases de Herbrand). Este artículo fue pionero en el desarrollo de lógicas probabilísticas en varios sentidos. Primero, propuso una interpretación geométrica para el conjunto de restricciones lineales inducido por el conjunto de sentencias y sus anotaciones probabilísticas. Esta interpretación claramente ilustra los métodos propuestos por Nilsson para calcular soluciones al problema intratable de obtener distribuciones de probabilidad que satisfagan las restricciones. Uno de dichos métodos para la inferencia probabilística consiste en aproximar el vector de variables asociado con la fórmula de consulta mediante una combinación lineal de las filas de la matriz que contiene los coeficientes de los mundos posibles, dependiendo de las relaciones lógicas entre las fórmulas en el conjunto de entrada y la consulta. El otro método propuesto por Nilsson consiste en calcular la distribución que tiene la *máxima entropía*; este enfoque fue adoptado más tarde por otros para su aplicación en la programación en lógica probabilística, como discutiremos más adelante.

Durante las últimas dos décadas del siglo 20 se continuó con el desarrollo de muchos trabajos en esta dirección. Halpern [14] considera dos formas de darle una semántica a la lógica probabilística de primer orden. La primera es adecuada para “conocimiento estadístico” o “escenarios de azar”, y asume la existencia de un mundo donde las afirmaciones acerca de individuos pueden ser verdaderas o falsas; de esta manera, no son representables los *grados de creencia*. El ejemplo usado por Halpern para ilustrar esta semántica es la frase “la probabilidad de que un pájaro elegido aleatoriamente vuele es mayor que 0,9”. El otro tipo de semántica asigna probabilidades a los mundos posibles, y, por lo tanto, no están bien definidos los escenarios de azar, y esto es más apropiado para grados de creencia. En relación con el ejemplo previo, un ejemplo de esta semántica es la frase “la probabilidad de que Tweety (un pájaro en particular) vuele

es mayor que 0,9”. Halpern prueba que ambos enfoques pueden ser combinados y aborda el problema de proveer axiomas sanos y completos para caracterizar el razonamiento probabilístico en los distintos tipos de escenario. Aunque hay resultados previos de complejidad [1] que muestran que no es posible una axiomatización *completa*, se establecen axiomas sanos para ambas semánticas que son suficientemente ricos como para realizar razonamiento probabilístico interesante. Se extrae además un paralelo entre esta situación y el resultado de incompletitud de Gödel para la axiomatización de la aritmética y los axiomas sanos (pero no completos) de Peano.

El trabajo de Jaumard et al. [16] es muy interesante ya que es posiblemente el primero en informar sobre implementaciones de la lógica probabilística. Los autores consideran extensiones de la lógica probabilística de Nilsson que involucran intervalos de valores de probabilidad, probabilidades condicionadas y un nuevo problema que implica el cálculo de mínimas modificaciones de anotaciones probabilísticas para restablecer la satisfacibilidad de un conjunto de sentencias. Este trabajo se centra en una propuesta basada en *generación de columnas* (*column generation*, en inglés) que permita resolver exactamente todas estas extensiones. La generación de columnas [11, 5] es una técnica diseñada para resolver eficientemente programas lineales con un gran número de variables dejándolas implícitas. Utiliza la teoría de la dualidad asignando restricciones a las variables y viceversa, y extiende el método revisado del símplice para determinar la columna entera resolviendo un subproblema auxiliar (también llamado el *oráculo*), o el *generador de columnas*. Encontrar una columna de este tipo enumerando todas las posibilidades es intratable, y este subproblema depende del tipo de problema considerado, pero es usualmente uno de programación combinatoria. Resolver el subproblema es por lo general NP-hard; sin embargo, no es necesario resolverlo exactamente en cada iteración, siempre y cuando se encuentre un costo reducido negativo. Cuando la heurística ya no genere tales valores, se debe utilizar un algoritmo exacto, lo cual es una debilidad del método. Uno de los resultados más importantes de [16] es que el *problema de la valoración* (encontrar las entradas para el oráculo) puede ser reducido a una instancia de MAX-SAT con pesos, que implica que PSAT puede ser resuelto en tiempo polinomial (por medio del método del elipsoide) para aquellas clases de fórmulas en las cuales MAX-SAT con pesos puede ser resuelto en tiempo polinomial.

Finalmente, el trabajo de Andersen y Pretolani [2] apunta a casos tratables de satisfacibilidad probabilística en fórmulas CNF cuyas cláusulas son anotadas con probabilidades. Los autores señalan que PSAT es computacionalmente más difícil que SAT, ya que se mantiene difícil en casos en los cuales SAT no lo hace. Su objetivo es entonces identificar casos fáciles de PSAT en los cuales es posible dar una representación compacta del conjunto de asignaciones probabilísticas consistentes. Se consideran dos enfoques diferentes basados en diferentes representaciones de fórmulas CNF. El primero está ba-

sado en hipergrafos dirigidos; extendiendo una formulación de programación entera de MAX-SAT, es posible resolver el caso en el cual el hipergrafo no tiene ciclos. Si una fórmula es representada por hiperárboles, su matriz asociada es equilibrada, lo cual permite proyectar restricciones en un espacio n -dimensional, dando lugar a un programa lineal pequeño. El otro enfoque está basado en grafos de co-ocurrencia, y los autores proveen una solución para el caso en el que el grafo es un 2-árbol parcial. En este caso, el resultado más importante afirma que PSAT puede ser reducido a resolver un sistema de $O(n)$ ecuaciones en $O(n)$ variables no negativas para fórmulas representables por 2-árboles parciales.

Programación en lógica probabilística. Los formalismos probabilísticos basados en programación en lógica fueron introducidos por primera vez por Ng y Subrahmanian en [26, 27]; estos dos artículos abordan el problema de combinar lógica probabilística [21] con teoría de la probabilidad, adoptando semánticas del estilo de Nilsson y Halpern como discutimos anteriormente. Los autores señalan que todas las semánticas propuestas para programación en lógica cuantitativa anteriores a su trabajo habían sido no probabilísticas, de las cuales son ejemplos [40] y [35]; este enfoque probabilístico tiene como objetivo el desarrollo de una teoría de modelos probabilísticos y una teoría de punto fijo. El tratamiento lógico de probabilidades en programación en lógica es complicado por dos razones: primero, los conectivos no pueden ser interpretados simplemente como funciones de valores de verdad tradicionales, y segundo, un conjunto de sentencias definidas con forma cercana a cláusulas y sin negaciones aún puede ser inconsistente. La forma general de las reglas en su formalismo es:

$$F_0 : \mu_0 \leftarrow F_1 : \mu_1 \wedge \dots \wedge F_n : \mu_n$$

donde los F_i son *fórmulas básicas* (conjunciones o disyunciones de átomos) y los μ_i son anotaciones probabilísticas en forma de intervalos que pueden contener expresiones con variables. En [27], las cabezas de las reglas están restringidas a átomos anotados, donde la negación se puede representar mediante anotaciones de la forma $[0, 0]$, pero las probabilidades condicionadas no pueden ser expresadas. Este formalismo es un marco lógico general para expresar información probabilística, y los autores estudian su semántica y su relación con la teoría de la probabilidad, la teoría de modelos, la teoría de punto fijo y la teoría de prueba. También desarrollan un procedimiento para responder consultas sobre probabilidades de eventos, que es diferente del procesamiento de consultas en programación en lógica clásica ya que los unificadores más generales no son siempre únicos y por lo tanto deben computarse unificadores maximalmente generales. En este formalismo, las fórmulas atómicas asignan rangos de probabilidad (intervalos) a los átomos y, en base a éstas, una función de fórmula determina el rango de probabilidad para una fórmula no atómica aplicando \oplus y \otimes (siguiendo el principio de la ignorancia). El operador de punto fijo se define con estas funciones de fórmulas.

La programación en lógica probabilística fue estudiada

más tarde por varios autores: Ngo y Haddawy [28], Lukasiewicz y Kern-Isberner [24], Lakshmanan y Shiri [20], Dekhtyar y Subrahmanian [7], Damasio et al. [6], entre otros. Ngo y Haddawy [28] presentan una teoría de modelos, una teoría de punto fijo y un procedimiento de prueba para la programación en lógica probabilística condicionada. Lukasiewicz y Kern-Isberner [24] combinan programación en lógica probabilística (adoptando también un tratamiento explícito de probabilidades condicionadas) con máxima entropía, como fue discutido más arriba en relación al propósito original de Nilsson. En un trabajo estrechamente relacionado, Lukasiewicz [22] presenta una semántica condicionada para programas lógicos probabilísticos donde cada regla se interpreta especificando la probabilidad condicionada de la cabeza de la regla, dado el cuerpo. Lakshmanan y Shiri [20] desarrollaron una semántica para programas lógicos en la cual los diferentes métodos axiomáticos generales son dados para calcular probabilidades de conjunciones y disyunciones, y estos son utilizados para definir una semántica para programas lógicos probabilísticos. En [7], Dekhtyar y Subrahmanian consideran diferentes estrategias para conjunciones y disyunciones, originalmente introducidas por Lakshmanan et al. en [19], y permiten una sintaxis explícita en programas lógicos probabilísticos para que los usuarios puedan expresar sus conocimientos de una dependencia. Damasio et al. [6] presentan una semántica bien fundada para programas lógicos anotados y muestran cómo calcularla.

Desarrollos más recientes. En las últimas dos décadas se ha visto una proliferación de desarrollos en el área de razonamiento probabilístico con modelos lógicos, tal vez por la combinación de dos factores: (i) el desarrollo de hardware rápido capaz de correr los algoritmos complejos asociados con este tipo de modelos, al menos para instancias de tamaño mediano; y (ii) la explosión en la cantidad de datos disponibles en la Web. La programación en lógica probabilística ha sido explorada en sus aplicaciones en el modelado de acciones de otros agentes [17, 37]; otra línea interesante es el desarrollo de métodos novedosos para razonar con probabilidades inexactas [4, 41], y la investigación de algoritmos para llevar a cabo abducción en programas lógicos probabilísticos [36]. Otra línea relevante para este tipo de programas lógicos es la del desarrollo de modelos para razonamiento temporal [33, 34].

También se han estudiado extensamente las bases de datos probabilísticas [9, 8, 38], en las cuales generalmente se hace una suposición de independencia probabilística entre tuplas. Cercanos a las bases de datos están los desarrollos de lenguajes ontológicos probabilísticos, en los cuales se estudia el uso de reglas al estilo de la programación en lógica pero buscando lograr una baja complejidad computacional; en esta línea se destacan las extensiones de Datalog+/- con anotaciones probabilísticas [12] y las lógicas de descripción probabilísticas [15, 18, 23, 25]. Por último, otros modelos relacionales probabilísticos tales como la Lógica de Markov [32] y PRL [10] han tenido un gran impacto en los desarrollos del área.

3. Líneas de Investigación, Desarrollo e Innovación

Dado que el objetivo de la línea de investigación es el de desarrollar un relevamiento comprensivo de la literatura sobre formalismos lógicos para el razonamiento probabilístico bajo incertidumbre, la metodología utilizada será la de búsqueda sistemática de bibliografía relacionada en las principales áreas en las que se han publicado trabajos pertinentes:

1. Bases de datos
2. Representación de conocimiento y razonamiento
3. Machine learning
4. Sistemas aplicados, investigando aplicaciones en la práctica e implementaciones disponibles.

En el último punto, se buscará realizar un relevamiento de las implementaciones existentes, su disponibilidad al público en general, y sacar conclusiones acerca de madurez (sistema prototipo de interés académico, sistema implementado en etapa beta, o sistema maduro). Esto ayudará a completar el panorama del estado actual del arte, y el resultado a su vez será una guía útil para todo investigador y practicante que esté interesado en utilizar estos formalismos.

4. Resultados y objetivos

El objetivo general de esta investigación es el estudio de los formalismos basados en diferentes formalismos lógicos que se han desarrollado para llevar a cabo razonamiento bajo incertidumbre probabilística. Como ya hemos visto, estos formalismos han sido estudiados en una gran variedad de áreas de las Ciencias de la Computación y la Matemática, y en la actualidad no hay literatura que unifique su tratamiento y discusión. El objetivo particular de esta tesis es entonces el de acortar las brechas entre las áreas y disciplinas que han abordado este tema, con el fin de evitar el desarrollo aislado de trabajos que comparten muchos objetivos.

5. Formación de Recursos Humanos

Esta línea de investigación corresponde a la tesis de Magister en Ciencias de la Computación de Natalia Abad Santos, bajo la dirección de Gerardo I. Simari, a desarrollarse dentro del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA) de la Universidad Nacional del Sur. Actualmente, el LIDIA cuenta con investigadores, becarios y estudiantes de posgrado trabajando intensamente en las áreas de Razonamiento bajo Incertidumbre e Inconsistencia, Web Semántica, Razonamiento sobre Preferencias, Robótica Cognitiva, Argumentación Rebatible, Revisión de Creencias y Sistemas

Multi-agente. El trabajo propuesto está estrechamente relacionado con varios proyectos que se están llevando a cabo en el lugar de trabajo propuesto. El director de tesis ha sido participante central en diferentes proyectos de investigación sobre lenguajes ontológicos para razonar bajo incertidumbre, razonamiento en redes complejas, sistemas multi-agente, revisión de creencias, robótica cognitiva, y razonamiento acerca del comportamiento de adversarios en el marco de proyectos de modelamiento cultural computacional.

Referencias

- [1] ABADI, M., AND HALPERN, J. Y. Decidability and expressiveness for first-order logics of probability. *Information and Computation* 112, 1 (1994), 1–36.
- [2] ANDERSEN, K. A., AND PRETOLANI, D. Easy cases of probabilistic satisfiability. *Ann. Math. Artif. Intell.* 33, 1 (2001), 69–91.
- [3] BOOLE, G. *The Laws of Thought*. Macmillan, London, 1854.
- [4] BROECHELER, M., SIMARI, G. I., AND SUBRAHMANIAN, V. Using histograms to better answer queries to probabilistic logic programs. In *Proc. of ICLP* (2009), pp. 40–54.
- [5] CHVÁTAL, V. *Linear Programming*. W.H.Freeman, New York, 1983.
- [6] DAMASIO, C. V., PEREIRA, L. M., AND SWIFT, T. Coherent well-founded annotated logic programs. In *Logic Programming and Non-monotonic Reasoning* (1999), pp. 262–276.
- [7] DEKHTYAR, A., AND SUBRAHMANIAN, V. S. Hybrid probabilistic programs. In *Proc. of ICLP* (1997), pp. 391–405.
- [8] DEY, D., AND SARKAR, S. A probabilistic relational model and algebra. *ACM Trans. on Database Systems* 21, 3 (1996), 339–369.
- [9] FUHR, N., AND RÖLLEKE, T. A probabilistic relational algebra for the integration of information retrieval and database systems. *ACM Trans. on Information Systems* 15, 1 (1997), 32–66.
- [10] GETOOR, L., AND GRANT, J. Prl: A probabilistic relational language. *Machine Learning* 62, 1-2 (2006), 7–31.
- [11] GILMORE, P. C., AND GOMORY, R. E. A linear programming approach to the cutting-stock problem. *Operations Research* 9, 6 (1961), 849–859.
- [12] GOTTLÖB, G., LUKASIEWICZ, T., MARTINEZ, M. V., AND SIMARI, G. I. Query answering under probabilistic uncertainty in Datalog+/- ontologies. *Ann. Math. Artif. Intell.* 69, 1 (2013), 37–72.

- [13] HALPERIN, T. Probability logic. *Notre Dame Journal of Formal Logic* 25 (3) (1984), 198–212.
- [14] HALPERN, J. Y. An analysis of first-order logics of probability. *Artif. Intell.* 46, 3 (Dec. 1990), 311–350.
- [15] HEINSOHN, J. Probabilistic description logics. In *Proc. of UAI* (1994), pp. 311–318.
- [16] JAUMARD, B., HANSEN, P., AND DE ARAGÃO, M. P. Column generation methods for probabilistic logic. In *Proc. of 1st. Integer Programming and Combinatorial Optimization Conference* (1990), pp. 313–331.
- [17] KHULLER, S., MARTINEZ, M. V., NAU, D., SIMARI, G., SLIVA, A., AND SUBRAHMANIAN, V. Computing most probable worlds of action probabilistic logic programs: Scalable estimation for $10^{30,000}$ worlds. *Ann. Math. Artif. Intell.* 51, 2–4 (2007), 295–331.
- [18] KOLLER, D., LEVY, A., AND PFEFFER, A. P-CLASSIC: a tractable probabilistic description logic. In *Proc. of AAAI* (1997), vol. 1997, pp. 390–397.
- [19] LAKSHMANAN, L. V. S., LEONE, N., ROSS, R., AND SUBRAHMANIAN, V. S. ProbView: a flexible probabilistic database system. *ACM Trans. on Database Systems* 22 (1997), 419–469.
- [20] LAKSHMANAN, L. V. S., AND SHIRI, N. A parametric approach to deductive databases with uncertainty. *IEEE Transactions on Knowledge and Data Engineering* 13, 4 (2001), 554–570.
- [21] LLOYD, J. W. *Foundations of Logic Programming, Second Edition*. Springer-Verlag, 1987.
- [22] LUKASIEWICZ, T. Probabilistic logic programming. In *Proc. of ECAI* (1998), pp. 388–392.
- [23] LUKASIEWICZ, T. Expressive probabilistic description logics. *Artif. Intell.* 172, 6 (2008), 852–883.
- [24] LUKASIEWICZ, T., AND KERN-ISBERNER, G. Probabilistic logic programming under maximum entropy. In *LNCS*, vol. 1638. 1999.
- [25] LUKASIEWICZ, T., MARTINEZ, M. V., ORSI, G., AND SIMARI, G. I. Heuristic ranking in tightly coupled probabilistic description logics. In *Proc. of UAI* (2012), pp. 554–563.
- [26] NG, R. T., AND SUBRAHMANIAN, V. S. A semantical framework for supporting subjective and conditional probabilities in deductive databases. In *Proc. of ICLP* (1991), The MIT Press, pp. 565–580.
- [27] NG, R. T., AND SUBRAHMANIAN, V. S. Probabilistic logic programming. *Information and Computation* 101, 2 (1992), 150–201.
- [28] NGO, L., AND HADDAWY, P. Probabilistic logic programming and bayesian networks. In *Asian Computing Science Conference* (1995), pp. 286–300.
- [29] NILSSON, N. Probabilistic logic. *Artif. Intell.* 28 (1986), 71–87.
- [30] PEIRCE, C. A theory of probable inference. In *Studies in Logic: By Members of Johns Hopkins University*, C. Peirce, Ed. Little, Brown and Company, Boston, 1883, pp. 126–181.
- [31] REICHENBACH, H. *The theory of probability, an inquiry into the logical and mathematical foundations of the calculus of probability*. English translation by Ernest H. Hutten and Maria Reichenbach, 2d ed. ed. University of California Press, Berkeley,, 1949.
- [32] RICHARDSON, M., AND DOMINGOS, P. Markov logic networks. *Machine learning* 62, 1-2 (2006), 107–136.
- [33] SHAKARIAN, P., PARKER, A., SIMARI, G. I., AND SUBRAHMANIAN, V. S. Annotated probabilistic temporal logic. *ACM Trans. Comput. Log.* 12, 2 (2011), 14.
- [34] SHAKARIAN, P., SIMARI, G. I., AND SUBRAHMANIAN, V. S. Annotated probabilistic temporal logic: Approximate fixpoint implementation. *ACM Trans. Comput. Log.* 13, 2 (2012), 13.
- [35] SHAPIRO, E. Y. Logic programs with uncertainties: A tool for implementing rule-based systems. In *Proc. of IJCAI* (1983), pp. 529–532.
- [36] SIMARI, G. I., DICKERSON, J. P., SLIVA, A., AND SUBRAHMANIAN, V. Parallel abductive query answering in probabilistic logic programs. *ACM Trans. Comput. Log.* 14, 2 (2013).
- [37] SIMARI, G. I., MARTINEZ, M. V., SLIVA, A., AND SUBRAHMANIAN, V. Focused most probable world computations in probabilistic logic programs. *Ann. Math. Artif. Intell.* 64, 2–3 (2012).
- [38] SUCIU, D., OLTEANU, D., RÉ, C., AND KOCH, C. *Probabilistic Databases*. Morgan & Claypool Publishers, 2011.
- [39] TARSKI, A. Wahrscheinlichkeitslehre und mehrwertige logik. *Erkenntnis* 5 (1935-36), 174–175.
- [40] VAN EMDEN, M. Quantitative deduction and its fixpoint theory. *Journal of Logic Programming* 4 (1986), 37–53.
- [41] YUE, A., LIU, W., AND HUNTER, A. Imprecise probabilistic query answering using measures of ignorance and degree of satisfaction. *Ann. Math. Artif. Intell.* 64, 2-3 (2012), 145–183.

Argumentación probabilística y revisión de creencias con aplicaciones a ciberseguridad

José N. Paredes Marcelo A. Falappa Gerardo I. Simari

Inst. de Cs. e Ing. de la Computación (Universidad Nacional del Sur-CONICET)
Av. Alem 1253, (8000) Bahía Blanca, Argentina
{jose.paredes,mfalappa,gis}@cs.uns.edu.ar

Resumen

Esta línea de investigación se centra en los aspectos algorítmicos y de representación de conocimiento y razonamiento asociados con los procesos de razonamiento dialéctico y dinámica del conocimiento bajo incertidumbre probabilística. La investigación es conducida por la aplicación de éstos en entornos relacionados con ciberseguridad y ciber guerra; dada la aplicabilidad de los resultados en datos provenientes del mundo real, la tratabilidad computacional es un tema central del proyecto.

Palabras clave: Representación de conocimiento y razonamiento, Argumentación, Revisión de creencias, Razonamiento bajo incertidumbre probabilística, Ciber seguridad.

1. Contexto

La presente línea de investigación se encuentra inserta en el marco del proyecto PGI 24/N035 “*Argumentación y Dinámica de Creencias para mejorar las capacidades de razonamiento y representación de conocimiento de Sistemas Multi-Agente*”, financiado por la Universidad Nacional del Sur, y el proyecto PIP-CONICET 112-201101-01000 “*Combinación de Revisión de Creencias y Argumentación para mejorar las capacidades de Razonamiento y modelado de la Dinámica de Conocimiento en Sistemas Multi-agente*”, financiado por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), ambos llevado a cabo dentro del Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur.

2. Introducción

A continuación haremos un breve relevamiento de las áreas más cercanas a esta línea de investigación, enfocándonos en los trabajos que comparten en mayor medida nuestros objetivos.

Sistemas argumentativos

Para los seres humanos, argumentar es una forma natural de encontrar una base segura para las creencias;

es una manera de manejar racionalmente la información contradictoria con el fin de establecer qué es posible creer en el contexto de lo que ha sido ya aceptado. La actividad humana de discutir, y la propia naturaleza de una discusión, han sido objeto de intensa investigación en Filosofía desde la antigüedad, y la Lógica nació del esfuerzo realizado para estructurar la presentación de argumentos y su intercambio. El proceso de argumentación (tanto la humana como la computacional) refleja una forma de razonamiento en el cual tanto la conclusión como la forma de llegar a ella pueden ser cuestionadas; es decir, se modela un proceso dialéctico en el cual las partes proponen argumentos a favor de sus puntos de vista o en contra de los de la contraparte, y el proceso determina cuál es la posición ganadora.

En la literatura se han propuesto muchas formas de llevar a cabo este proceso, las cuales se dividen principalmente en abstractas [Dun95] y concretas [GS04]; las primeras se caracterizan por el análisis de los argumentos disponibles y la relación de ataque entre ellos, mientras que las últimas asumen la disponibilidad de la estructura interna de cada argumento. Esta área de estudio es de especial interés en el ámbito de la Inteligencia Artificial, principalmente porque permite razonar con información incompleta e incierta, y permite manejar inconsistencias en los sistemas basados en conocimiento. En las últimas dos décadas, la investigación en argumentación como mecanismo de razonamiento se ha expandido enormemente, y el campo está produciendo un gran variedad de resultados.

Es claro que este tipo de razonamiento puede ser útil para nuestros objetivos; como punto de partida para esta investigación se considerarán tanto los sistemas de argumentación abstracta como los sistemas concretos basados en reglas. En particular, se tomará como referente a tres sistemas concretos que han sido desarrollados en la última década: Defeasible Logic Programming (DeLP) [GS04], ABA [DKT09] y ASPIC+ [Pra10], pero también se tendrá en cuenta la aproximación abstracta, ya que en general resulta útil en las etapas iniciales de modelado cuando es necesario establecer qué propiedades debe cumplir el sistema.

Razonamiento bajo incertidumbre probabilística

Las herramientas para llevar a cabo un proceso de razonamiento probabilístico han sido estudiadas por mucho tiempo y en muchas disciplinas, dado su amplio campo de aplicación. Para nuestros objetivos, utilizaremos herramientas maduras que poseen implementaciones optimizadas y mantenidas periódicamente; dos ejemplos son las Redes Bayesianas [Pea14] y las Redes Lógicas de Markov [RD06], aunque se realizará un relevamiento completo del campo para identificar otras posibilidades.

Revisión de creencias

El problema que se estudia en el área de Revisión de Creencias (*Belief Revision*, en inglés) es el de cómo deben cambiar los estados epistémicos de un agente ante una nueva entrada epistémica; en otras palabras, cómo deben revisarse las creencias ante la presencia de información nueva que posiblemente contradice las creencias establecidas hasta el momento. Tradicionalmente, los estados epistémicos han tomado la forma o bien de conjuntos de creencias (conjuntos de fórmulas lógicas cerrados bajo consecuencia) [AGM85, Gär03] o bases de creencias (conjuntos no cerrados) [Han94, Han97].

Es evidente que el problema de la revisión de creencias aparece constantemente en el mundo real y, por lo tanto, cualquier sistema inteligente que funcione con datos del mismo deberá estar equipado con alguna forma de revisión. La metodología típica consiste en el desarrollo de operadores que toman la base de conocimiento actual y la entrada epistémica, y producen una nueva base de conocimiento que corresponde al resultado de revisar las creencias. Estos operadores en general se caracterizan por las propiedades que deben cumplir (expresadas en la forma de postulados); luego se proponen construcciones algorítmicas y se demuestra formalmente que las dos caracterizaciones son equivalentes; éste tipo de resultado lleva el nombre de teorema de representación.

De particular interés para esta línea de investigación es la capacidad de hacer revisión de creencias en entornos en los que se espera que haya intenciones de engañar al agente razonador (con información engañosa); por ejemplo, una persona o equipo que ejecuta un ciberataque típicamente intenta dejar pistas falsas para que los investigadores pierdan el rastro del verdadero responsable. A nuestro entender, no existen operadores de revisión de creencias que se hayan desarrollado con esta característica en mente. Por otra parte, también se explorarán operadores que sean especialmente propicios para funcionar con aspectos cuantitativos; el único trabajo en esta línea es la propuesta reciente de [SSF15].

3. Líneas de Investigación, Desarrollo e Innovación

Los últimos años han visto una verdadera explosión en el interés por el desarrollo y estudio de formalismos de

argumentación probabilística. Si bien la mayor parte de esta atención ha estado enfocada hasta el momento en sistemas de argumentación abstracta [Hun12, Thi12], existen también desarrollos con sistemas argumentativos concretos; el primer trabajo en esta línea fue el de [HKL00]; desde entonces, los trabajos más relevantes para este plan han sido los de Hunter [Hun13] y Shakarian, Simari y Falappa [SSF14].

La conexión entre la revisión de creencias y los sistemas argumentativos fue estudiada por primera vez en [Doy79]; en [FGKIS11] se realiza un relevamiento hasta la fecha de publicación y además se desarrollan operadores nuevos que explotan esta conexión. Los operadores basados en explicaciones de [FKIS02] son también relevantes a la combinación de estas dos áreas, dado que las explicaciones de por qué una revisión se hace de una manera particular pueden ser vistas como argumentos.

Por otra parte, la revisión de creencias probabilísticas también ha recibido algo de atención en la literatura, aunque menos que las discutidas anteriormente. Los trabajos funcionales en la línea de investigación más relevante para la nuestra — la que combina métodos de revisión clásicos con conocimiento probabilístico — son los de [KIR04], que se centra en la fusión de información probabilística, y [CD05], que utiliza además evidencia incierta como entrada epistémica. Más recientemente, en [Dub11] se propone un tratamiento general que abarca la revisión y fusión en entornos tanto cualitativos como cuantitativos.

Por último, la única línea de investigación de la que tenemos conocimiento que combina las tres áreas es la iniciada recientemente por Shakarian, Simari, Falappa y otros [SSM⁺15, SSF14, SSMP15, SSF15]. Estos trabajos plantean un formalismo general para realizar revisión de creencias en un lenguaje basado en DeLP extendido con presuposiciones [MGS12] y anotaciones probabilísticas. El modelo consiste de una base de conocimiento bipartita: el *modelo analítico* es un programa DeLP que incluye los datos y reglas que aplican al dominio en cuestión, mientras que el *modelo del entorno* es una base de conocimiento probabilística. Por último, una *función de etiquetado* relaciona ambas partes al asociar fórmulas compuestas por combinaciones de eventos básicos provenientes del modelo del entorno. En la Figura 1 se muestra un ejemplo de predicados que podrían utilizarse para modelar un dominio de ciberseguridad; una función de etiquetado podría por ejemplo asociar *motivo*(X, Y) con el evento probabilístico *enConf*(X, Y), denotando que se puede concluir que X tiene motivo de atacar a Y cuando éstos están en conflicto entre ellos.

Si bien los autores estudian operadores de revisión para este modelo, no han estudiado formas de hacerlos computacionalmente tratables, no se han enfocado en la revisión de información potencialmente producida con la intención de engañar, ni han estudiado su aplicación a datos provenientes del mundo real. En nuestra línea de investigación trabajaremos con este grupo en estos aspectos, no sólo enfocados en sus desarrollos anteriores, sino abriendo también la búsqueda a otros formalismos.

\mathbf{P}_{EM} :	$origIP(M, X)$	El malware M originó de una dirección de IP perteneciente a X .
	$malwOp(M, O)$	El malware M fue usado en la ciber-operación O .
	$malwPista(M, X)$	El malware M contiene una pista de que fue creado por X .
	$idiomaComp(M, I)$	El malware M fue compilado en un sistema que usa el idioma I .
	$idiomaNat(X, I)$	I es el idioma nativo de X .
	$enConf(X, Y)$	X e Y están en conflicto entre ellos.
	$numInstMCI(X, N)$	En el país X hay al menos N instituciones de primer nivel en matemática, ciencias e ingeniería.
	$infSisGob(X, M)$	Sistemas de gobierno de X fueron infectados con el malware M .
	$edadCapCib(X, N)$	X has tenido capacidades de ciber-guerra por N años o menos.
	$capCibGob(X)$	X tiene un laboratorio gubernamental de ciber-seguridad.
\mathbf{P}_{AM} :	$condOp(X, O)$	X condujo la ciber-operación O .
	$evidencia(X, O)$	Hay evidencia de que X condujo la ciber-operación O .
	$motivo(X, Y)$	X tiene un motivo para lanzar un ciber-ataque en contra de Y .
	$esCapaz(X, O)$	X es capaz de conducir una ciber-operación O .
	$objetivo(X, O)$	X fue el objetivo de la ciber-operación O .
	$invMCI(X)$	X tiene inversiones significativas en educación en matemática, ciencia e ingeniería.
	$expCib(X)$	X tiene experiencia en la conducción de ciber-operaciones.

Figura 1: Un ejemplo muy sencillo de un conjunto de predicados que podrían utilizarse para modelar un dominio de ciber-guerra. \mathbf{P}_{EM} corresponden al modelo de entorno (probabilístico), mientras que \mathbf{P}_{AM} contiene los predicados del modelo analítico (de argumentación).

4. Resultados y objetivos

El objetivo general de este plan de trabajo es investigar los aspectos tanto de representación de conocimiento como algorítmicos asociados con los procesos de razonamiento dialéctico y dinámica del conocimiento bajo incertidumbre probabilística. El enfoque particular será en su aplicación en entornos relacionados con ciberseguridad y ciberguerra; dado que se quieren aplicar los resultados en datos provenientes del mundo real, la tratabilidad computacional deberá ser tenida en cuenta a lo largo de todo el proyecto. Uno de los problemas principales que ocurre en ciberseguridad y ciberguerra es el llamado “problema de la atribución”, que busca encontrar la parte culpable de un acto de interés en el ciberespacio (como, por ejemplo, un acceso no autorizado a una base de datos) [Tsa12, SSR13].

El objetivo particular posee dos partes principales:

1. El desarrollo de herramientas que combinen formalismos de argumentación estructurada con modelos probabilísticos, con especial enfoque en la capacidad de considerar múltiples fuentes de información que pueden contener información elaborada por enemigos con la meta de engañar al agente razonador; de ahora en más, nos referiremos a ésta como “información engañosa” (*deceptive information*, en inglés). Para ello, se evaluarán – entre otras opciones – diferentes aproximaciones a la revisión de creencias no priorizada. Este objetivo incluye el análisis de las propiedades teóricas de los resultados obtenidos, tales como su complejidad computacional en tiempo y espacio.
2. La evaluación experimental de las herramientas obtenidas en el objetivo (a); habrá dos enfoques particulares: (i) su capacidad para representar información que ocurre en escenarios del mundo real; y (ii)

escalabilidad: que el sistema sea capaz de funcionar con una gran cantidad de información en un tiempo razonable; dado que la meta es resolver problemas complejos (tales como la decisión de quienes son responsables por un ataque); un “tiempo razonable”, en este contexto, puede ser unas pocas horas. Resaltamos la diferencia entre este tipo de problemas y otros que suelen requerir una respuesta casi inmediata, como una consulta a una base de datos. Los dos objetivos particulares se influenciarán entre sí; el primero se enfoca en los aspectos teóricos, mientras que el segundo se encarga de probar cómo funcionan los desarrollos teóricos en la práctica.

Los pasos iniciales en este proyecto involucran el estudio profundo de sistemas argumentativos y modelos probabilísticos, que son las componentes básicas de los formalismos a desarrollar. A su vez, se estudian las diferentes formas de hacer revisión de creencias, identificando aproximaciones innovadoras que sean propicias para: (i) el modelado de problemas que incluyen información engañosa, y (ii) entornos en los cuales es necesario razonar con incertidumbre probabilística.

5. Formación de Recursos Humanos

Dentro de esta línea de investigación se lleva a cabo la tesis de Doctor en Ciencias de la Computación de José Paredes, bajo la dirección de Marcelo A. Falappa y Gerardo I. Simari, en desarrollo dentro del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA) de la Universidad Nacional del Sur. Actualmente, el LIDIA cuenta con investigadores, becarios y estudiantes de posgrado trabajando intensamente en las áreas de Razonamiento bajo Incertidumbre e Inconsistencia, Web

Semántica, Razonamiento sobre Preferencias, Robótica Cognitiva, Argumentación Rebatible, Revisión de Creencias y Sistemas Multi-agente.

Los directores propuestos han llevado a cabo diferentes proyectos de investigación sobre Sistemas Argumentativos Basados en Reglas y su aplicación en áreas tales como Toma de Decisiones, Robótica Cognitiva y Web Semántica (entre otras). Dentro de estos proyectos se desarrolló una implementación conocida como DeLP (por sus siglas en Inglés de *Defeasible Logic Programming*)[GS04] y hoy cuenta con una versión disponible que permite el uso del sistema a través de la Web¹. Se ha desarrollado también una versión llamada DeLP-Server que provee un servicio de razonamiento para sistemas multi-agentes donde diferentes agentes en host remotos pueden hacer uso de este servicio de razonamiento argumentativo [GRTS07]. Por último, el director propuesto posee vínculos estrechos con el grupo de investigación del Dr. Paulo Shakarian (Arizona State University, EE.UU.), experto no sólo en temas de ciberseguridad y ciberguerra (es ex-oficial del ejército estadounidense, incluyendo tareas de analista de inteligencia), sino también en formalismos de razonamiento bajo incertidumbre en general.

Referencias

- [AGM85] Carlos E Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The journal of symbolic logic*, 50(02):510–530, 1985.
- [CD05] Hei Chan and Adnan Darwiche. On the revision of probabilistic beliefs using uncertain evidence. *Artificial Intelligence*, 163(1):67–90, 2005.
- [DKT09] Phan Minh Dung, Robert A Kowalski, and Francesca Toni. Assumption-based argumentation. In *Argumentation in artificial intelligence*, pages 199–218. Springer, 2009.
- [Doy79] Jon Doyle. A truth maintenance system. *Artificial intelligence*, 12(3):231–272, 1979.
- [Dub11] Didier Dubois. Information fusion and revision in qualitative and quantitative settings. In *Proc. of ECSQARU*, pages 1–18. Springer, 2011.
- [Dun95] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n -person games. *Artificial intelligence*, 77(2):321–357, 1995.
- [FGKIS11] Marcelo A Falappa, Alejandro J García, Gabriele Kern-Isberner, and Guillermo R Simari. On the evolving relation between belief revision and argumentation. *The Knowledge Engineering Review*, 26(01):35–43, 2011.
- [FKIS02] Marcelo A Falappa, Gabriele Kern-Isberner, and Guillermo R Simari. Explanations, belief revision and defeasible reasoning. *Artificial Intelligence*, 141(1):1–28, 2002.
- [Gär03] Peter Gärdenfors. *Belief revision*, volume 29. Cambridge University Press, 2003.
- [GRTS07] Alejandro J García, Nicolás D Rotstein, Mariano Tucac, and Guillermo R Simari. An argumentative reasoning service for deliberative agents. In *Knowledge Science, Engineering and Management*, pages 128–139. Springer, 2007.
- [GS04] Alejandro J García and Guillermo R Simari. Defeasible logic programming: An argumentative approach. *Theory and practice of logic programming*, 4(1+ 2):95–138, 2004.
- [Han94] Sven Ove Hansson. Kernel contraction. *The Journal of Symbolic Logic*, 59(03):845–859, 1994.
- [Han97] Sven Hansson. Semi-revision. *Journal of Applied Non-Classical Logics*, 7(1-2):151–175, 1997.
- [HKL00] Rolf Haenni, Jürg Kohlas, and Norbert Lehmann. *Probabilistic argumentation systems*. Springer, 2000.
- [Hun12] Anthony Hunter. Some foundations for probabilistic abstract argumentation. pages 117–128, 2012.
- [Hun13] Anthony Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54(1):47–81, 2013.
- [KIR04] Gabriele Kern-Isberner and Wilhelm Rödder. Belief revision and information fusion on optimum entropy. *International Journal of Intelligent Systems*, 19(9):837–857, 2004.
- [MGS12] Maria Vanina Martinez, Alejandro Javier García, and Guillermo Ricardo Simari. On the use of presumptions in structured defeasible reasoning. In *Proc. of COMMA 2012*, pages 185–196, 2012.
- [Pea14] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2014.
- [Pra10] Henry Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.

¹http://lidia.cs.uns.edu.ar/delp_client/

- [RD06] Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine learning*, 62(1-2):107–136, 2006.
- [SSF14] Paulo Shakarian, Gerardo I Simari, and Marcelo A Falappa. Belief revision in structured probabilistic argumentation. In *Proc. of FoIKS*, pages 324–343. Springer, 2014.
- [SSF15] Gerardo I. Simari, Paulo Shakarian, and Marcelo A. Falappa. A quantitative approach to belief revision in structured probabilistic argumentation. *Annals of Mathematics and Artificial Intelligence*, pages 1–34, 2015.
- [SSM⁺15] Paulo Shakarian, Gerardo I. Simari, Geoffrey Moores, Damon Paulo, Simon Parsons, Marcelo A. Falappa, and Ashkan Aleali. Belief revision in structured probabilistic argumentation: Model and application to cyber security. *Annals of Mathematics and Artificial Intelligence*, pages 1–43, 2015.
- [SSMP15] Paulo Shakarian, Gerardo I Simari, Geoffrey Moores, and Simon Parsons. Cyber attribution: An argumentation-based approach. In *Cyber Warfare*, pages 151–171. Springer, 2015.
- [SSR13] Paulo Shakarian, Jana Shakarian, and Andrew Ruef. *Introduction to cyber-warfare: A multidisciplinary approach*. Newnes, 2013.
- [Thi12] Matthias Thimm. A probabilistic semantics for abstract argumentation. In *Proc. of ECAI*, pages 750–755, 2012.
- [Tsa12] Nicholas Tsagourias. Cyber attacks, self-defence and the problem of attribution. *Journal of Conflict and Security Law*, pages 229–244, 2012.

Modelos de Análisis de Información para la Toma de Decisiones Estratégicas del Sector Tealero

Karina B. Eckert^a, Fabián E. Favret^b, Matias M. Barboza^c, Leandro M. Witzke^d, Victor M. Alvarenga^e

Universidad Gastón Dachary

Avda. López y Planes 6519, Posadas, Misiones-Argentina. Tel: +54 (0376) – 447699

{ karinaeck^a, fabianfavret^b, matias.mbz^c, leanwitzke^d }@gmail.com, alva_victor@hotmail.com^e

Resumen

Para mantenerse persistente en el mercado, las organizaciones deben estar atentas a la evolución continúa de los factores externos. Esto le permite generar estrategias que la vuelvan competitiva o, al menos, generar estrategias correctivas que le permitan seguir funcionando. En ese contexto, determinar qué información es útil o no, requiere de un proceso de análisis exhaustivo, que se enfrenta con la sobresaturación de información disponible actualmente.

Hoy en día existen métodos y herramientas que permiten a las organizaciones hacer lo que se conoce como Vigilancia Tecnológica (VT) e Inteligencia Competitiva (IC). Estos procesos tienen como objetivo estar alertas a los cambios constantes de la tecnología y el mercado, a fin de poder ajustar las estrategias de la organización. Ambos involucran procesos inteligentes e intentan detectar y predecir los cambios del entorno a fin de que la organización pueda articular mecanismos de anticipación. Este proyecto tiene como objetivo implementar técnicas de VT e IC utilizando algoritmos de minería de la WEB para examinar documentos de dominio público y obtener, a partir de ellas, información de valor para la toma de decisiones referido a la cadena de

producción y comercialización del té; sin embargo la utilidad del modelo trasciende la temática donde es aplicado.

Palabras clave: Toma de Decisiones Estratégicas, Vigilancia Tecnológica e Inteligencia Competitiva, Minería WEB, Cadena de Producción y Comercialización del Té.

Contexto

Este trabajo de investigación se encuentra en ejecución en el marco del 6° Llamado a Concurso de Proyectos de Investigación, convocado por la Secretaría de Investigación y Desarrollo de la Universidad Gastón Dachary en el año 2015. En el proyecto se radican dos Trabajos Finales de Carrera (TFC) en la UGD y una Tesis de Maestría de Administración de Negocios en la UTN-FRRe.

Introducción

Vigilancia Tecnológica e Inteligencia Competitiva

Actualmente existen dos actividades que las empresas usan cada vez más, para analizar las variaciones de su entorno: la Vigilancia Tecnológica (VT) y la Inteligencia Competitiva (IC). Ambas tienen el objetivo de alertar cambios en las tendencias en la evolución de la tecnología y el

comportamiento del mercado. La VT incluye todos aquellos métodos que intentan anticipar y entender la dirección potencial, las características y efectos de los cambios tecnológicos, especialmente en lo referido a invención innovación y uso [1]. La VT es un proceso continuo de captación selectiva de información del exterior y de la propia organización referida a ciencia y tecnología, que se transforma luego en conocimiento para minimizar los riesgos en el proceso de toma de decisiones y anticipar los cambios del entorno.

La IC es un conjunto de métodos éticos y legales que una organización puede utilizar a fin de poder obtener información valiosa sobre sus competidores. Básicamente la IC tiene como objetivo conseguir información sobre las actividades de su entorno y anticipar el comportamiento futuro de los competidores, proveedores, clientes, mercados, productos y servicios [2]. No se intenta obtener información confidencial de los competidores sino simplemente captar, analizar y procesar información de dominio público que está disponible y al alcance de cualquier persona.

Claramente, ambas estrategias de captación de datos deben integrarse en sistemas que den apoyo a la toma de decisiones que analicen la información obtenida. Esas aplicaciones se conocen como Sistemas de Soporte de Decisiones (SSD) y proveen una serie de técnicas y herramientas que sirven para asistir al decisor durante el proceso de toma de decisiones [3][4][5].

La gran evolución de las TICs ha provocado la generación y el fácil acceso a una extraordinaria cantidad de información. En este contexto, una de las áreas en las que se estudian y desarrollan métodos para buscar información de valor en grandes volúmenes de datos se conoce como Minería de Datos (DM) [6]; enfo-

cada en el desarrollo de algoritmos que permitan extraer conocimiento desde los datos.

Desde el punto de vista de la VT y la IC, existe una actividad de gran utilidad derivada de DM: el análisis de la información de la WEB. Estos algoritmos se utilizan básicamente para obtener información relevante de grandes volúmenes en contenido WEB.

Básicamente los algoritmos de minería WEB (WM) es el uso de técnicas de DM para descubrir y extraer información de los documentos y recursos disponibles de la WEB [7].

Esta actividad puede clasificarse en tres categorías: minería del contenido WEB; minería de la estructura de la WEB; y minería del uso de la WEB [8]. De estas tres categorías, la más relevante para la VT y la IC es la minería del contenido WEB, la cual abarca el descubrimiento de recursos [9], categorización y clustering de documentos [10] y extracción de información de páginas WEB [11].

Los algoritmos de WM proveen las herramientas adecuadas para monitorear los cambios del entorno de la organización. Debido a que el dinamismo propio de Internet aporta complejidad al problema de la búsqueda de información, se necesitan técnicas inteligentes adaptables a la variabilidad de las condiciones de la WEB.

Evidentemente, contar con algoritmos confiables y eficientes para la exploración de textos y de la WEB permite determinar el valor correcto que tiene la información analizada para una organización. Por lo tanto, se justifica el desarrollo modelos de análisis de información que utilicen técnicas WM para VT e IC.

Cadena de Producción y Comercialización del Té

En el mercado del té, Argentina ocupa un lugar significativo en el segmento del té negro a granel (2,7%), mientras que la participación en el conjunto de las exportaciones de té es mucho menor (0,5%). Comparando la posición relativa en orden mundial de las exportaciones, ocupa el puesto 42 en el mundo cuando se considera la totalidad de productos relacionados con el té, mientras que en té negro a granel ocupa el sexto lugar a nivel internacional. En el té verde a granel ocupa el puesto 19, mientras en los té fraccionados están ubicados en rangos inferiores al número 43. Estos datos revelan una debilidad en la incorporación de valor agregado en el té argentino [12].

La producción tealera se concentra en la provincia de Misiones corresponde al 95 % de la superficie de té cultivada en el país; el otro 5% corresponde a el noreste de la provincia de Corrientes [12].

En el año 2006, se conformó el Agglomerado Productivo (AP) del sector Tealero, que involucra a actores de distintos eslabones de la cadena: empresas privadas y cooperativas integrantes del sector tealero; APAM (Asociación de Productores Agropecuarios de Misiones); la Comisión Provincial del Té (COPROTÉ); el Instituto Nacional de Tecnología Agropecuaria - Estación Experimental Agropecuaria de Cerro Azul (INTA-EEA Cerro Azul); la Universidad Nacional de Misiones (UNaM); la Universidad Gastón Dachary (UGD); el Ministerio del Agro y la Producción de la provincia de Misiones; la Municipalidad de Oberá.

La cadena de valor del AP tealero está integrada por cuatro eslabones:

- Producción primaria de té. Los productores agrícolas realizan las acciones de plantación, desmalezamiento, cosecha y transporte de brotes de té.
- Primera transformación. Comprende básicamente el marchitado, enulado,

fermentado y secado de las hojas y brotes de té.

- Segunda transformación. Comprende el despalado, desfibrado, tipificación, y envasado en diversas formas.
- Comercialización. Dirigida al mercado interno con diversas modalidades de marcas, envases, y tipos de té; y al mercado externo a granel en bolsas de papel multipliego con una lámina interior de aluminio.

Entre los beneficios que se podrían obtener a través de la correcta obtención de información del sector, se pueden mencionar los siguientes:

- Ampliar y recuperar mercados. Analizar industrias como la alemana, que solo le dan valor agregado al té, no siembran ni hacen cosecha.
- Analizar nuevos mercados con necesidades de valor agregado en el rubro del té.
- Armar y conocer a fondo la cadena de valor de proveedores (transporte, químicos, bolsas, etc.).
- Conocer sobre los precios del mercado al instante, por tipos de té (hoja, palo o mix). Encontrar información sobre Tecnologías de plantación (viveros por ejemplo).
- Detectar alternativas para las épocas fuera de cosecha.

Líneas de Investigación, Desarrollo e Innovación

Este proyecto implica el desarrollo de cinco etapas para el desarrollo del modelo de análisis de información basado en técnicas de minería de WEB:

1. La primera etapa consiste en el estudio y análisis de las metodologías de Vigilancia Tecnológica e Inteligencia Competitiva. Búsqueda, análisis y resumen de las técnicas y herramientas actualmente utilizadas para realizar VT e

IC. Entrevistas con los especialistas a fin de obtener información de calidad y actualizada sobre el área.

2. La segunda etapa abarca el estudio y análisis de los requerimientos de información de la cadena de producción y comercialización del té (productor, elaborador, exportador, consumidor). El estudio incluye la identificación y caracterización de fuentes de información relevante en el dominio de aplicación.

3. La tercera etapa implica el estudio, análisis y desarrollo de algoritmos de WM. Relevamiento bibliográfico correspondiente utilizando las fuentes disponibles de información a fin de establecer qué desarrollos han tenido éxito en el área. Programación de los algoritmos que se consideren relevantes, con las modificaciones necesarias para realizar VT e IC.

4. La cuarta etapa comprende el estudio y análisis de los Sistemas de Soporte de Decisiones Estratégicas. Elevamiento de los modelos que se consideren adecuados para realizar VT e IC proponiendo el que mejor se adecúe a las necesidades del proyecto.

5. Finalmente la etapa cinco incluye la integración de las técnicas propuestas. Análisis, diseño y desarrollo de las técnicas estudiadas, utilizando como campo de aplicación el cluster del té.

Actualmente se está trabajando el desarrollo de una herramienta basada en técnicas de WM que sirva de soporte a la toma de decisiones en el ámbito del clúster del té en Misiones. Se ha realizado un estudio del campo confeccionando un listado de las necesidades de información con la que cuenta el clúster y se han estudiado los algoritmos de WM que puedan ser utilizados para buscar y capturar contenido de documentos relevantes en Internet. También se han analizado las diferentes metodologías de ranking para determinar la relevancia de cada docu-

mento WEB en función a su contenido. El trabajo abarca el análisis, diseño, desarrollo del prototipo y la validación de sus resultados. El prototipo de software propuesto es implementado en lenguaje Python, utilizando la librería Pattern, respetando un diseño orientado a objetos. Se abarca el estudio de distintos métodos de ponderación de documentos WEB en base a su contenido, seleccionando aquellos que favorezcan al desarrollo de una herramienta eficiente para la toma de decisiones del clúster. Una vez obtenida la información considerada relevante, la misma será presentada en una página HTML. Las técnicas de recolección de información utilizadas son los algoritmos de búsqueda automática de información (WEB “crawling”), análisis de relevancia de documentos, ranking y extracción de contenido (WEB “scraping”). Los resultados finales serán evaluados por miembros del clúster mediante encuestas con una metodología de valoración adjunta.

Resultados y Objetivos

Objetivo general

Desarrollar un modelo de análisis de información disponible en la WEB (referida a tecnología y competidores) que sea útil para el proceso de toma de decisiones estratégicas en la cadena de producción y comercialización del té.

Objetivos Específicos

- Revisar las alternativas disponibles de Sistemas de Soporte de Decisiones estratégicas.
- Estudiar y Analizar las herramientas y métodos de VT e IC para la obtención y el análisis de información.
- Analizar e implementar los algoritmos y técnicas de análisis de información utilizando WM.

- Integrar las técnicas de WM orientando su utilización práctica con información de la producción tealera de la Provincia de Misiones.

Formación de Recursos Humanos

El equipo del trabajo de investigación está compuesto por docentes y alumnos en instancias de TFC del Departamento de Ingeniería y Ciencias de la Producción de la UGD y UTN-FRRe.

Director del Proyecto:

Dr. Marcelo Karanik

Co-Director:

Ing. Roberto Suénaga

Docentes-Investigadores:

Ing. Karina Eckert

Ing. Fabián Favret

Lic. Tokuji Kairiyama

Tesis de Maestría de Administración de Negocios en UTN-FRRe:

Ing. Jorge Roa

Alumnos en instancia TFC:

Alvarenga Victor

Barboza Matias

Witzke Leandro

Referencias

- [1] A. Firat, W. Woon, and S. Madnick, "Technological Forecasting – A Review," presented at the Working Paper CISL# 2008-15, Cambridge, 2008.
- [2] R. G. Vedder, M. T. Vanecek, C. Guynes, and J. Cappel, "CEO and CIO Perspectives on Competitive Intelligence," *Commun. ACM*, vol. 42, no. 8, 1999.
- [3] J. P. Shim, M. Warkentin, J. F. Courtney, D. J. Power, R. Sharda, and C. Carlsson, "Past, Present, and Future of Decision Support Technology," *Decis. Support Syst.*, vol. 33, no. 2, pp. 111–126, 2002.
- [4] C. W. Holsapple, "Decision Support Systems," *Encycl. Inf. Syst.*, vol. 1, pp. 551–565, 2003.
- [5] D. J. Power and R. Sharda, "Decision support systems," *Springer Handbook of Automation*, pp. 1539–1548, 2009.
- [6] S. H. Weiss and M. Indurkha, *Predictive Data Mining: A Practical Guide*. San Francisco, CA: Morgan Kaufmann Publishers, 1998.
- [7] O. Etzioni, "The World Wide WEB: Quagmire or gold mine," *Commun. ACM*, vol. 39, no. 11, pp. 65–68, 1996.
- [8] R. Kosala and H. Blokeel, "WEB mining research: A survey," *ACM SIGKDD Explor.*, vol. 2, no. 1, pp. 1–15, 2000.
- [9] S. Chakrabarti, M. van der Berg, and B. Dom, "Focused crawling: A new approach to topic-specific WEB resource discovery," in *In Proceedings of the 8th international conference on World Wide WEB (WWW8)*, 1999, pp. 545–562.
- [10] T. Kohonen, S. Kaski, K. Lagus, J. Salojarvi, V. Paatero, and A. Saarela, "Selforganization of a massive document collection," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 574–585, 2000.
- [11] C. H. Chang, C. N. Hsu, and S. C. Lui, "Automatic information extraction from semi-structured WEB pages by pattern discovery," *Decis. Support Syst.*, vol. 35, pp. 129–147, 2003.
- [12] Ministerio de Economía y Finanzas Públicas, "Programa de Competitividad del Norte Grande. 'Plan de Competitividad Conglomerado Tealero, Provincia de Misiones.'" [Online]. Available: http://www.mecon.gov.ar/programanorte grande/docs/plan_misiones_tealero_actualizacion.pdf. [Accessed: 15-Apr-2015].

Optimización de rutas en el transporte de personas para la pequeña y mediana empresa

Villagra A., Villagra S., Alancay N., Rasjido J., Pandolfi D.

Laboratorio de Tecnologías Emergentes (LabTEm)

Instituto de Tecnología Aplicada - Unidad Académica Caleta Olivia

Universidad Nacional de la Patagonia Austral

{avillagra, svillagra, nalancay, jrasjido, dpandolfi}@uaco.unpa.edu.ar

Resumen

El transporte juega un papel importante en las tareas de logística de muchas compañías, ya que normalmente representa un alto porcentaje del valor añadido a los bienes. Por tanto, la utilización de métodos computacionales en el transporte suele producir ahorros sobre su costo total. Varias compañías de transporte utilizan métodos manuales, a veces basados en la experiencia de los expertos, para la planificación de sus operaciones. En otros casos se utilizan reglas heurísticas para mejorar la planificación manual. Sin embargo, solo el uso de modernas técnicas de optimización permite abordar problemas de alta complejidad.

El diseño y optimización de rutas utilizando computación inteligente es ventajoso en cualquier ámbito y situación para cualquier tipo de usuario, en especial para las empresas de transporte cuyas pérdidas y ganancias se basan en la distribución óptima tanto del tiempo, como del combustible, que están directamente relacionadas con la distancia recorrida.

En el presente trabajo se describe la investigación abordada en el campo de las metaheurísticas para resolver este problema principalmente aplicando algoritmos híbridos basados en el algoritmo genético celular (cGA) y en la optimización basada en colonia de hormigas (*Ant Colony Optimization* o sus siglas en inglés ACO).

Palabras clave: Algoritmos Genéticos Celulares, Hibridación, Metaheurísticas, Optimización Basada en Colonia de hormigas, Problemas de Ruteo de Vehículos.

Contexto

La línea de investigación presentada en este documento se lleva a cabo en el Laboratorio de Tecnologías Emergentes (LabTEm) de la Universidad Nacional de la Patagonia Austral, Unidad Académica Caleta Olivia. En el marco del programa del Ministerio de Educación Secretaría de Políticas Universitarias (SPU) en el “Programa Universidad y Transporte” a través del Proyecto de I&D denominado “Optimización de rutas de transporte de personas para la pequeña y mediana empresa”. Código 32-64-122. Resolución Nro. 3332/15

Introducción

En la vida cotidiana existe la necesidad de transportar personas u objetos de un lugar a otro. Dentro de los procesos requeridos para suplir esta necesidad se encuentra el de definir las rutas que deben realizar los vehículos. Las decisiones tomadas en la definición de las rutas pueden implicar un gran ahorro o desperdicio de recursos en la labor del transporte, por lo cual, este proceso tiene una gran relevancia en la cadena de abastecimiento. Algunas de las situaciones de la vida cotidiana en las que se requiere planear y programar el ruteo de

vehículos son las siguientes: empresas productoras, empresas de transporte de bienes, mensajería, transporte de desechos, empresas de transporte de personas, transporte escolar, empresas de transporte de valores, etc.

Existen varios métodos para la optimización de redes logísticas centradas en el transporte urbano de personas y objetos como por ejemplo: TSP (*Traveling salesman problem*), CPP (*Chinese postman problem*) y finalmente el VRP (*Vehicle routing problem*). En esta línea de investigación nos centraremos en este último.

El VRP consiste en generar rutas de reparto dado una cantidad de clientes por atender, un conjunto de vehículos de reparto y un punto de origen, permitiendo minimizar ciertos factores que ayuden a la empresa a obtener beneficios [6][14].

El interés de este problema viene dado por dos causas principales. Por un lado, el VRP es un problema *NP-duro* [15] y de alto interés académico debido a su dificultad en las restricciones que incluye, y en la multitud de variantes existentes. Por otro lado, muchos problemas del mundo real pueden ser visualizados (o concebidos) como variantes de VRP. Existe una evolución constante en la calidad de las metodologías empleadas para resolverlo, tanto algoritmos exactos como métodos heurísticos (secuenciales y paralelos). Debido a la dificultad que presenta el problema, no existe ningún método exacto capaz de resolver instancias de más de 50 clientes [20].

Algunas de las metaheurísticas más comúnmente utilizadas en el VRP y sus variantes son por ejemplo, los Algoritmos Genéticos (AGs) [10] que han tenido éxito en resolver problemas de ruteo de vehículos, corte de empaquetado (*Strip Packing*), entre muchos otros.

Líneas de investigación y desarrollo

En esta sección se describe la línea de investigación que se lleva a cabo en el proyecto.

El concepto de optimización puede verse como el proceso de encontrar y mejorar el rendimiento de una aplicación o dispositivo a partir de determinados cambios lógicos o físicos.

Las técnicas aproximadas, llamadas metaheurísticas, consisten básicamente en la combinación de métodos heurísticos básicos en plataformas de más alto nivel con el fin de explorar el espacio de búsqueda de una forma eficiente y efectiva. En [5] se pueden encontrar recopiladas varias definiciones de metaheurísticas dadas por diferentes autores. Entre algunas metaheurísticas podemos nombrar: *Simulated Annealing* [12], *Scatter Search* (SS) [9], procedimiento de búsqueda adaptativo aleatoriamente voraz (*Greedy Randomized Adaptive Search Procedure-GRASP*) [17], búsqueda por vecindario variable (*Variable Neighborhood Search-VNS*) [11], y otros basados en una población, tales como algoritmos evolutivos (AEs) [3], optimización por colonia de hormigas [7], optimización por cúmulo de partículas (*Particle Swarm Optimization-PSO*) [8], las cuales se abordarán en esta línea de investigación.

En los últimos años se han obtenido buenos resultados en muchos problemas de optimización clásicos y de la vida real utilizando metaheurísticas híbridas. Talbi en [18] y [19] propone una taxonomía para algoritmos híbridos y presenta dos clasificaciones para este tipo de algoritmos: jerarquizada y plana. Esta clasificación establece esquemas híbridos específicos en donde, en general, diferentes algoritmos son combinados de acuerdo a ciertos criterios. Sin embargo, en esta línea se pretende

brindar una perspectiva diferente al esquema híbrido planteado previamente. Más precisamente, se pretende crear algoritmos cuya construcción siga criterios similares a los establecidos para diseñar algoritmos híbridos, pero incorporando componentes de los algoritmos antes que al algoritmo como un todo. Tomando en una primera etapa como anfitriones a cGA y ACO.

En cuanto a las metaheurísticas celulares [1] se basan en el concepto de vecindario. La exploración y la difusión de las soluciones, al resto de la población, se produce debido a que los vecindarios están solapados, lo que produce que las buenas soluciones se extiendan lentamente por toda la población. Aparte de estos modelos básicos, en la literatura también se han propuesto modelos híbridos donde se implementan esquemas de dos niveles.

En lo relacionado a ACO aplicado a VRP y sus variantes existen diversos trabajos como por ejemplo [2, 4, 13 y 23]. Se analizarán las propuestas con el objetivo detectar fortalezas y debilidades para proponer mejoras.

Por lo tanto, es un desafío la exploración de algunas variantes de problema, la búsqueda de metodologías y estrategias de solución.

Resultados obtenidos/esperados

Durante los últimos años el grupo ha analizado, hibridado y comparado distintas metaheurísticas para resolver el problema de VRP. Utilizamos un algoritmo genético simple AG, el algoritmo MCMP-SRI como algoritmo base, y propuestas híbridas [16 y 21].

Hemos realizado un estudio sobre un mecanismo de hibridación basado en el concepto de componentes activas que definimos como la “esencia” de una metaheurística. En otras palabras, aquellas

partes de una metaheurística que caracterizan su comportamiento en cuanto a la forma que exploran el espacio de búsqueda. Además, hemos definido una metodología para su identificación y su posterior aplicación en una metaheurística anfitriona, en nuestro caso el cGA [22]. Hemos trabajado con PSO, SA y SS identificando componentes activas utilizando la metodología propuesta. Teniendo en cuenta los resultados obtenidos, podemos decir que los algoritmos híbridos propuestos lograron un alto porcentaje de éxito en la obtención del valor óptimo. Estos resultados nos alientan a expandir el conjunto de problemas, en particular a variantes de VRP. Además de extender y mejorar la metodología para identificar componentes activas y aplicarlas a diferentes metaheurísticas anfitrionas.

El proyecto tiene como objetivo optimizar en tiempo y distancia el recorrido de las rutas de transporte a través de la utilización de diferentes técnicas de inteligencia computacional que impacta directamente sobre los costos de las pymes.

Particularmente, el desarrollo de una herramienta para la gestión de rutas, mediante algoritmos metaheurísticos híbridos, para obtener un recorrido óptimo de transporte de personas para pymes de la zona Norte de la provincia de Santa Cruz.

Formación de recursos humanos

Un integrante de este proyecto de investigación está desarrollando su Tesis de Doctorado en temáticas afines.

Un integrante ha comenzado su Maestría orientando sus cursos y trabajos a esta línea de investigación.

Se cuenta con un becario de investigación de grado.

Referencias

- [1] Alba E. y Dorronsoro B. *Cellular Genetic Algorithms*. Springer 2008.
- [2] Aziz E., An Algorithm for the Vehicle Problem, *International Journal of Advanced Robotic Systems*, Vol. 7, No. 2, 2010, pp. 125-132.
- [3] Bäck T., Fogel D., y Michalewicz Z., editors. *Handbook of Evolutionary Computation*. Oxford University Press, 1997.
- [4] Balseiro S., Loiseau I. and Ramone J., An Ant Colony Algorithm Hybridized with Insertion Heuristics for the Time Dependent Vehicle Routing Problem with Time Windows, *Computers & Operations Research*, 2011, pp. 954-966. doi:10.1016/j.cor.2010.10.011.
- [5] Blum C. y Roli A. Metaheuristics in combinatorial optimization: Overview y conceptual comparison. *ACM Computing Surveys*, 35(3):268–308, 2003.
- [6] Christofides, N., Mingozzi A. y Toth P. The vehicle routing problem. *Reveu française d'automatique d'informatique et de recherche opérationnelle. Journal Combinatorial Optimization*. 1:315–338, 1979.
- [7] Corne D., Dorigo M., Glover F., editors. The ant colony optimization metaheuristic. *New Ideas in Optimization*, pages 11–32. McGraw Hill, 1999.
- [8] Eberhart R. y Kennedy J. A new optimizer using particles swarm theory. In *Sixth International Symposium on Micro Machine y Human Science (Nagoya, Japan)*, IEEE Service Center, Piscataway, pages 39–43, 1995.
- [9] Glover F. Heuristics for integer programming using surrogate constraints. *Decision Sciences*, 8:156-166, 1977.
- [10] Goldberg, D. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley. 1989.
- [11] Hansen P. y Mladenovic N. Variable neighborhood search for the p-median. *Location Science*, 5(4):207–226, 1997.
- [12] Kirkpatrick S, Gelatt J, Vecchi M. Optimization by Simulated Annealing. *Science* 220:671–680, 1983.
- [13] Kumar S. and Panneerselvam R. A survey on the vehicle routing problem and its variants. *Intelligent Information Management*, 4(3) 66-74 <http://dx.doi.org/10.4236/iim.2012.43010> Published Online May 2012.
- [14] Laporte, G. The vehicle routing problem: An overview of exact and approximate algorithms. *European J. Operational Research*. 59:345-358, 1992.
- [15] Lenstra J.K. y Rinnooy Kan A.H.G., “Complexity of vehicle routing y scheduling problems,” *Networks*. 11:221–227, 1981.
- [16] Mercado V., Villagra A., Pandolfi D., Leguizamón G. Algoritmos Evolutivos multirecombinativos híbridos aplicados al problema de ruteo de vehículos con capacidad limitada. *XIX Congreso Argentino de Ciencias de la Computación*. 2013. Mar del Plata. ISBN 978-987-23963-1-2.
- [17] Resende M. y Ribeiro C. *Handbook of Metaheuristics*, chapter Greedy randomized adaptive search procedures, pages 219–249. Kluwer Academic Publishers, 2003.
- [18] Talbi, E.-G. A taxonomy of hybrid metaheuristics. *Heuristics, Journal of heuristics*, 8(5):541–564, 2002.
- [19] Talbi, E.-G. *Metaheuristics: From design to Implementation*. Wiley, 2009.
- [20] Toth P. y Vigo D., *The Vehicle Routing Problem*, Monographs on Discrete Mathematics y Applications. SIAM, Philadelphia, 2001.
- [21] Villagra A., Pandolfi D., Rasjido J., Mercado V. Hibridación de Metaheurísticas aplicadas al problema de ruteo de vehículos con capacidad uniforme. *IX Seminario Euro Latinoamericano de Sistemas de Ingeniería*. Venezuela. Noviembre 2013.

[22] Villagra A., Leguizamón G., and Alba E. Active components of metaheuristic in cellular genetic algorithms. *Soft Computing*, pages 1-15, 2014.

[23] Yu B. and Yang Z. An Ant Colony Optimization Model: the Period Vehicle Routing Problem with Time Windows. *Transportation Research Part E*, Vol. 47, 2011, pp. 166-181.

Busqueda local iterada para resolver problemas de planificación

Carlos Bermudez, Gabriela Minetti, Carolina Salto

Departamento de Informática/ Facultad de Ingeniería/ Universidad Nacional de
La Pampa

Calle 110 N°390, General Pico, La Pampa
{bermudezc, minettig, saltoc}@ing.unlpam.edu.ar

Resumen

El diseño de técnicas y algoritmos eficientes que resuelvan adecuadamente problemas complejos de optimización es uno de los campos dentro de la investigación en Informática con mayor repercusión en la actualidad. Por tanto, se han estudiado, diseñado y desarrollado un conjunto heterogéneo de metaheurísticas para resolver importantes problemas de optimización en el campo de la ingeniería.

Por tal motivo nuestra propuesta es proporcionar herramientas de software que permitan resolver problemas comunes en la mayoría de las empresas de la región en el sector productivo y en la logística de las mismas.

Palabras clave: Metaheurísticas, optimización combinatoria, planificación de tareas, búsqueda local iterada

Contexto

Esta línea de investigación se desarrolla en el marco del proyecto de investigación dirigido por la Dra. Gabriela Minetti (acreditado por la Facultad de Ingeniería – UNLPam) y del

PICTO0278-UNLPam (acreditado por la Agencia Nacional de Promoción Científica y Tecnológica). Ambos llevados a cabo en el Laboratorio de Investigación de Sistemas Inteligentes (LISI), de la Facultad de Ingeniería de la Universidad Nacional de La Pampa (entidad que acredita). Cabe destacar que desde hace varios años, los integrantes de estos proyectos mantienen una importante vinculación con investigadores de la Universidad Nacional de San Luis (Argentina) y de la Universidad de Málaga (España), con quienes se realizan publicaciones conjuntas.

Introducción

La búsqueda local iterada (Iterated Local Search, ILS) [1] es una metaheurística basada en un concepto simple pero muy efectivo. En cada iteración, la solución actual es perturbada y a esta nueva solución se le aplica un método de búsqueda local para mejorarla. Este nuevo mínimo local obtenido por el método de mejora puede ser aceptado como nueva solución actual si pasa un test de aceptación. La importancia del proceso de perturbación es obvia: si es demasiado pequeño puede que el

algoritmo no sea capaz de escapar del mínimo local; por otro lado, si es demasiado grande, la perturbación puede hacer que el algoritmo sea como un método de búsqueda local con un reinicio aleatorio. Por lo tanto, el método de perturbación debe generar una nueva solución que sirva como inicio a la búsqueda local, pero que no debe estar muy lejos de la actual para que no sea una solución aleatoria. El criterio de aceptación actúa como contrabalance, ya que filtra la aceptación de nuevas soluciones dependiendo de la historia de búsqueda y de las características del nuevo mínimo local.

Este algoritmo es relativamente simple de implementar y usa muy pocos parámetros, lo cual facilita el proceso de ajuste de los mismos. Por lo tanto, ILS se transforma en un algoritmo atractivo para dar solución a problemas del mundo real. El método ha sido aplicado exitosamente a varios problemas de optimización combinatorial, entre los que se puede mencionar: planificación jobshop [1] [2] [3], problema de flowshop [3], ruteo de vehículos [4], MAXSAT [6].

El poder potencial de la búsqueda local iterada reside en su muestreo sesgado del conjunto de óptimos locales. La eficiencia de este muestreo depende tanto de los tipos de perturbaciones como de los criterios de aceptación. Pero todavía se pueden obtener mucho mejores resultados si se optimizan los módulos de búsqueda local iteradas.

El objetivo que se persigue con el desarrollo de esta línea de investigación está relacionada con la optimización del ILS para resolver problemas de optimización relativos a la producción y la logística, en particular a problemas de programación de tareas en ambientes industriales, tales como flowshop y jobshop.

Líneas de Investigación, Desarrollo e Innovación

En esta línea de investigación se considera una variante al jobshop clásico, donde cada operación se puede realizar en una máquina elegida de un conjunto finito de máquinas elegibles. El objetivo es hallar una asignación para cada operación en una máquina apropiada (problema de ruteo) y definir una secuencia de operaciones en cada máquina que minimice el tiempo necesario para terminar todas las tareas (*makespan*). A esta variante se la conoce como problema de jobshop flexible (*flexible jobshop problem* FJSP)

Para este problema, se desarrollan variantes del ILS, donde se analizan varios operadores para generar la perturbación de la solución actual y distintas búsquedas locales, a fin de determinar cuál de las combinaciones permite una mejor exploración del espacio de soluciones del FJSP. Como operadores de perturbación y de búsqueda se consideran operadores que no requieren mucho cómputo como lo son el operador de inserción y el de intercambio para permutaciones. La diferencia entre el operador de perturbación y el de búsqueda local, es que considerando la solución actual, el primero realiza un solo movimiento (intercambio o inserción según corresponda), mientras que el segundo considera todos los posibles movimientos.

Además del FJSP, en esta línea de investigación se trata una variante del flowshop donde existen varias etapas en la producción con múltiples máquinas (idénticas o no) por etapa, además cabe la posibilidad de que algún trabajo no

necesite ser procesado en algunas de las etapas. Esta variante se conoce como problema de flowshop flexible híbrido (*hybrid flexible flowshop problem*, HFFSP). Además, en varias industrias (farmacéutica, metalúrgica, automotriz, entre otras) se consideran tiempos de puesta a punto (*setup*) de las máquinas entre dos trabajos diferentes, los cuales aportan mayor dificultad a los problemas de planificación de la producción. Estos tiempos de setup pueden o no depender de la secuencia. Considerando estas restricciones, el problema considerado en esta línea de investigación es el HSSFP con tiempos de setup dependientes de la secuencia (*sequence dependent setup time*, SDTS/FSSFP), cuyo objetivo es minimizar el tiempo de finalización.

Se analiza el comportamiento del ILS para el SDST/HFFSP considerando distintos operadores de búsqueda local, proponiéndose una mejora al algoritmo propuesto por Naderi et al. [7]. Este último utiliza una búsqueda local (Local Search, LS) basada en el operador de inserción. Con el objetivo de incrementar la calidad de las soluciones, en esta línea de investigación se propone una variante de la metaheurística propuesta por Naderi, denominada ILSint. Dicha variante consiste en modificar el proceso de búsqueda local. Esta nueva LS se basa en el operador de intercambio.

Con el fin de evaluar nuestro aporte, estas propuestas algorítmicas se comparan con algoritmos que representen el estado del arte para ambos problemas (SDST/HFFS y FJSP).

Desarrollo

Los resultados obtenidos usando ILS para resolver tanto el FJSP como el SDST/HFFS [9] son prometedores.

En el caso del ISL para resolver el FJSP, los resultados experimentales sobre varias instancias de un conocido caso de prueba [8] verifican la efectividad de este algoritmo cuando se lo compara con otros algoritmos de la literatura [9]. Es importante destacar que la versión de ILS propuesta obtuvo los mejores resultados empleando el menor esfuerzo computacional en un alto porcentaje de los casos de prueba.

Del análisis de los resultados obtenidos del SDST/HFFS, se desprendió que la mejora realizada al ILS, ILSint, usando el operador de intercambio realiza una explotación mucho más eficaz del espacio de búsqueda que al aplicar el operador de inserción (propuesta original de [10]). Este algoritmo se compara con distintas versiones del algoritmo de recocido simulado, determinándose que los algoritmos basados en ILS obtuvieron soluciones de mejor calidad que los basados en recocido simulado y el esfuerzo computacional para lograrlo fue significativamente menor.

Formación de Recursos Humanos

Cada año se incorporan al proyecto alumnos avanzados en la carrera Ingeniería en Sistemas, quienes trabajan en temas relacionados a la resolución de problemas de optimización usando técnicas inteligentes, con el objeto de guiarlos en el desarrollo de sus tesis de grado y, también, de formar futuros investigadores científicos. Por otra parte,

los docentes-investigadores que integran el proyecto realizan diversos cursos de posgrado relacionados con la temática del proyecto, con el objetivo de sumar los créditos necesarios para cursar carreras de posgrado.

Referencias

- [1] H. Lourenço, O. Martin y T. Stützle., «Iterated local search,» de *Handbook of Metaheuristics*, Kluwer Academic Publishers, 2002, p. 321–353.
- [2] H. Lourenço, «Job-shop scheduling: Computational study of local search and large-step optimization methods,» *European Journal of Operational Research*, vol. 83, pp. 347-364, 1995.
- [3] H. Lourenço y M. Zwijnenburg, «Combining the large-step optimization with tabu-search: Application to the job-shop scheduling problem,» de *Meta-Heuristics: Theory & Applications.*, Kluwer Academic Publishers, 1996, pp. 219-236.
- [4] S. Kreipl, «A large step random walk for minimizing total weighted tardiness in a job shop,» *Journal of Scheduling*, vol. 3, nº 3, p. 125–138, 2000.
- [5] A. Juan, H. Lourenço, M. Mateo y Q. Castella, «Using iterated local search for solving the flow-shop problem: Parallelization, parametrization, and randomization issues,» *International Transactions in Operational Research*, vol. 21, nº 1, pp. 103-126, 2014.
- [6] P. Penna, A. Subramanian y L. Ochi, «An Iterated Local Search Heuristic for the Heterogeneous Fleet Vehicle Routing Problem,» *Journal of Heuristics*, vol. 19, nº 2, pp. 201-213, 2013.
- [7] A. Roli, «Design of a New Metaheuristic for MAXSAT Problems,» de *Principles and Practice of Constraint Programming - CP 2002*, Volume 2470 of the series Lecture Notes in Computer Science, 2002, pp. 767-767.
- [8] B. Naderi, R. Ruiz y M. Zandieh, «Algorithms for a realistic variant of flowshop scheduling,» *Computers & Operations Research*, vol. 2, nº 236-246, p. 37, 2010.
- [9] P. Brandimarte, «Routing and Scheduling in a Flexible Job-shop by Tabu Search,» *Annals of Operations Research*, vol. 22, nº 2, pp. 157-183, 1993.
- [10] M. González, C. Vela y R. Varela, «An Efficient Memetic Algorithm for the Flexible Job Shop with Setup Times,» de *Proceedings of the Twenty-Third International Conference on Automated Planning and Scheduling*, 2013.
- [11] B. Naderi, R. Ruiz y M. Zandieh, «Algorithms for a realistic variant of flowshop scheduling,» *Computers & Operations Research*, vol. 37, nº 2, pp. 236-246, 2010.

Tratamiento Masivo de Datos Utilizando Técnicas de Machine Learning

Claudia Russo¹, Hugo Ramón¹, Nicolás Alonso¹, Benjamin Cicerchia², Leonardo Esnaola¹,
Juan Pablo Tessore²

¹Instituto de Investigación y Transferencia en Tecnología (ITT) / Escuela de
Tecnología/ Universidad Nacional del Noroeste de la Provincia de Buenos Aires
(UNNOBA)
Sarmiento y Newbery, 236-4636945/44

²Becario de la Comisión de Investigaciones Científicas de la Provincia de Buenos
Aires (CIC)

claudia.russo@itt.unnoba.edu.ar / hugo.ramon@itt.unnoba.edu.ar /
nicolas.alonso@itt.unnoba.edu.ar / lucas.cicerchia@itt.unnoba.edu.ar /
leonardo.esnaola@itt.unnoba.edu.ar /
juanpablo.tessore@itt.unnoba.edu.ar

Resumen

Machine Learning es un área de la inteligencia artificial que engloba un conjunto de técnicas que hacen posible el aprendizaje automático a través del entrenamiento con grandes volúmenes de datos. Hoy en día existen diferentes modelos que utilizan esta técnica y consiguen una precisión incluso superior a la de los humanos en las mismas tareas, por ejemplo en el reconocimiento de objetos en una imagen.

La construcción de modelos de *Machine Learning* requiere adaptaciones propias debido a la naturaleza de los datos o a la problemática a la que se aplica. Así, surge la necesidad de investigar las diferentes técnicas que permitan obtener resultados precisos y confiables en un tiempo razonable.

Palabras clave:

Machine Learning, Big Data, Sistemas Inteligentes.

Contexto

Esta línea de investigación forma parte del proyecto “Tecnologías exponenciales en contextos de realidades mixtas e interfaces avanzadas.” aprobado por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA en el marco de la convocatoria a Subsidios de Investigación Bianuales (SIB2015). A su vez se enmarca en el contexto de planes de trabajo aprobados por la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires y por la Secretaría de Investigación de la UNNOBA en el marco de la convocatoria “Becas de Estudio Cofinanciadas 2015 CIC Universidades del interior bonaerense”.

El proyecto se desarrolla en el Instituto de Investigación en Tecnologías y Transferencia (ITT) dependiente de la

mencionada Secretaría, y se trabaja en conjunto con la Escuela de Tecnología de la UNNOBA.

El equipo está constituido por docentes e investigadores pertenecientes al ITT y a otros Institutos de Investigación, así como también, estudiantes de las carreras de Informática de la Escuela de Tecnología de la UNNOBA.

Introducción

Desde que las primeras computadoras programables fueron concebidas, las personas se preguntaron si tendrían la capacidad de pensar, de aprender y de convertirse en “máquinas inteligentes”.

El campo de la ciencia que se encarga de resolver este interrogante se denomina inteligencia artificial. Se trata de un área multidisciplinaria, que a través de ciencias como las ciencias de la computación, la matemática, la lógica y la filosofía, estudia la creación y diseño de sistemas capaces de resolver problemas cotidianos por sí mismos, utilizando como paradigma la inteligencia humana [1]. Para que una máquina pueda comportarse de manera inteligente debería ser capaz de resolver problemas de la manera en que lo hacen los humanos, es decir, en base a la experiencia y el conocimiento [2]. Esto implica que debería ser capaz de modificar su comportamiento en base a cuan precisos son los resultados obtenidos comparados con los esperados.

En este sentido podemos encontrar tres grandes grupos de algoritmos de *Machine Learning* [3]:

- Algoritmos supervisados: estos algoritmos utilizan un conjunto de datos de entrenamiento etiquetados (preclasificados), los cuales procesan para realizar predicciones sobre los mismos, corrigiéndolas cuando son incorrectas. El proceso de entrenamiento continúa hasta que el modelo alcanza un nivel deseado de precisión.
- Algoritmos semi-supervisados: combinan tanto datos etiquetados como no etiquetados para generar una función deseada o clasificador. Este tipo de modelos deben aprender las estructuras para organizar los datos así como también realizar predicciones.
- Algoritmos no supervisados: El conjunto de datos no se encuentra etiquetado y no se tiene un resultado conocido. Por ello deben deducir las estructuras presentes en los datos de entrada, lo puede conseguir a través de un proceso matemático para reducir la redundancia sistemáticamente u organizando los datos por similitud.

Dentro de esta clasificación podemos además encontrar un gran número de algoritmos específicos con diferentes características para el tratamiento de los datos. Entre los más relevantes encontramos:

- *Deep Learning*: consiste en la utilización de algoritmos para hacer representaciones abstractas de la información y facilitar el aprendizaje automático [4].
- *Active Learning*: es un caso especial de aprendizaje semi-supervisado donde el algoritmo de aprendizaje

puede interactuar con un usuario u otra fuente de información para obtener los resultados deseados [5].

- *Support Vector Machines*: busca la maximización de la distancia entre la recta o el plano y las muestras que se encuentran a un lado u otro. En el caso que las muestras no sean linealmente separables se utiliza una transformación llamada *kernel* [6] [7].

Líneas de Investigación, Desarrollo e Innovación

La presente investigación se encuadra dentro del eje “Tratamiento masivo de datos” y su procesamiento a través de sistemas inteligentes. En este sentido se pretende procesar señales provenientes de fuentes diversas, según la problemática investigada, para construir los conjuntos de entrenamiento necesarios. Así como también, la selección, diseño y desarrollo de un modelo que utilice alguno de los algoritmos relevados para lograr una correcta clasificación y predicción.

Se deberán abarcar las siguientes cuestiones:

- Obtención de un conjunto de datos suficientemente representativo para la problemática que se desea abordar y su clasificación.
- Pre procesamiento de las señales para lograr su normalización y adecuación.
- Resolver cuestiones relacionadas con el procesamiento de datos en tiempo

real y optimización de su funcionamiento.

- Análisis y selección de los distintos algoritmos de *Machine Learning* apropiados para el tipo de señal recibida (imágenes, video, sonido o incluso texto) y para la problemática de su aplicación.
- Evaluación de fiabilidad y desempeño de las diferentes técnicas de *Machine Learning* aplicadas.

Resultados y Objetivos

Se espera que la presente línea de I/D permita adquirir conocimientos específicos sobre las diferentes técnicas de *Machine Learning*, con el propósito de desarrollar modelos capaces de predecir y clasificar las señales involucradas en la problemática que se intenta resolver, obteniendo un comportamiento inteligente de manera automática.

Debido a que la universidad se encuentra dentro de la pampa húmeda, una de las regiones más relevantes en lo que respecta a producción agrícola, se pretende combinar agricultura de precisión con técnicas de machine learning y remote sensing [8] con el objetivo de dar soporte a la toma de decisiones en el sector.

Por otro lado, debido a las necesidades de los municipios de la región, se vislumbra la posibilidad de trabajar en la prevención y la detección de diferentes tipos de comportamiento, problemática que también puede ser atacada con este tipo de técnicas.

También se prevé la aplicación de machine learning en el análisis del texto en publicaciones periodísticas, contenido en foros y redes sociales, con la finalidad de encontrar patrones dentro de esos datos que permitan predecir comportamientos futuros en ámbitos específicos.

Así mismo, se busca generar informes técnicos en base al trabajo realizado, en donde se registren los avances, el grado de implementación y los resultados obtenidos. Como así también difundir y transferir los resultados y logros alcanzados mediante la presentación y participación en diferentes congresos, jornadas y workshops de carácter nacional e internacional vinculados a la temática de estudio.

Formación de Recursos Humanos

En esta línea de I/D se han obtenido y se encuentran desarrollando actualmente dos Becas de Estudio Cofinanciadas otorgadas por la Comisión de Investigaciones Científicas (CIC) y la UNNOBA. Asimismo se espera desarrollar cuatro tesis doctorales y dos tesinas de grado, dirigidas por miembros de este proyecto.

Bibliografía

- [1] Assessment of the Commercial Applicability of Artificial Intelligence in Electronic Businesses. Thomas Kramer. Diplom.de. 2002.
- [2] Data Classification Algorithms and Applications, Charu C. Aggarwal, CRC Press, 2015.
- [3] Machine Learning An Algorithmic Perspective Second Edition, Stephen Marsland, CRC Press, 2015.
- [4] A Deep Learning. Book in preparation for MIT Press. Bengio, Y., Goodfellow, I. and Courville, USA, 2015.
- [5] Active Learning Literature Survey, Settles Burr, Computer Sciences Technical Report 1648. University of Wisconsin–Madison, 2014.
- [6] A Tutorial on Support Vector Machines for Pattern Recognition, Christopher J.C. Burges, Kluwer Academic Publishers, 1998.
- [7] Top 10 algorithms in data mining, Xindong Wu et al. Knowledge and Information Systems 2008.
- [8] Machine learning in remote sensing data processing, Gustavo Camps-Valls, IEEE International Workshop on Machine Learning for Signal Processing, 2009.

APLICACIÓN DE MINERÍA DE DATOS ESPACIAL EN EL AREA DE SALUD EN LA ZONA DE INFLUENCIA DE LA UNNOBA.

Claudia Russo¹, Javier Charme¹, María Rosana Piergallini¹, Ma. Mercedes Guasch¹, Adriana Torriggino¹, Ana Smail¹

Instituto de Investigación y Transferencia en Tecnología (ITT), Escuela de Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)

Sarmiento y Newbery (CP 6000), Junín, Buenos Aires, Argentina. Teléfonos (0236) 4636945/44

{claudia.russo, javier.charme, rosana.piergallini, mercedes.guasch, adriana.torriggino, ana.smail }@itt.unnoba.edu.ar

Resumen

El aumento del volumen y variedad de información que se encuentra digitalizada en base de datos de diversas organizaciones ha crecido exponencialmente en las últimas décadas.

Sobre esta información es posible aplicar técnicas que nos permiten extraer conocimiento útil desde los datos almacenados y que se engloban bajo la denominación de Knowledge Discovery in Databases (KDD), donde Minería de Datos (MD) es un paso del KDD.

Cuando en la base de datos se incorpora el dato espacial, dada la complejidad de estos tipos de datos, los objetos que se almacenan (puntos, líneas, polígonos), y su estructura de datos se dificulta la utilización de la minería de datos tradicional. La Minería de Datos Espacial (MDE) provee un gran grupo de técnicas y herramientas para la

explotación de estos datos que permiten encontrar patrones potencialmente útiles.

El número de base de datos espaciales está creciendo constantemente y la diversidad de áreas en que este tipo de bases de datos es utilizada es muy amplio: medicina, geología, química, astronomía solo por mencionar algunos de ellas.

En salud y epidemiología se valora la importancia de la componente espacial en sus investigaciones y en el diseño de estrategias diferenciadas de prevención y control. En salud pública, el uso de las tecnologías de la información ha facilitado el conocimiento de los nuevos problemas de salud como la prevención de posibles eventos futuros.

A través de esta línea de investigación se propone aplicar técnicas de MDE para la extracción de nuevos conocimientos que asistan a la toma de decisiones en Salud Pública en el área de influencia de la Universidad Nacional del Noroeste de la Provincia de Buenos Aires

(UNNOBA). Los avances en este sentido contribuirán a preservar y mejorar el nivel de la salud pública en general.

Palabras clave: KDD, Minería de Datos, Minería de Datos Espacial, Salud Publica

Contexto

La línea de trabajo que aquí se presenta se inscribe en el marco del Proyecto de Investigación “Tecnologías exponenciales en contextos de realidades mixtas e interfaces avanzadas”, aprobado por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA, convocatoria a Subsidios de Investigación Bianuales (SIB2015).

El tema de estudio de nuestra línea de trabajo es la utilización de MDE en salud pública y epidemiología en el área de influencia de la Universidad Nacional del Noroeste de la Provincia de Buenos Aires.

El proyecto se desarrolla en el Instituto de Investigación y Transferencia en Tecnología (ITT), en la sede Pergamino dependiente de la mencionada Secretaría, y se trabaja en conjunto con la Escuela de Tecnología y el Departamento de Salud de la UNNOBA.

El equipo está constituido por docentes e investigadores que desarrollan actividades en el departamento de Informática y Tecnología y en el departamento de Salud de la UNNOBA.

Introducción

Las necesidades de toma de decisiones requieren el descubrimiento de nuevos modelos no esperados o imposibles de descubrir manualmente a partir de un gran volumen de datos.

El KDD es el “Proceso no trivial de identificar patrones válidos, novedosos,

potencialmente útiles y comprensibles a partir de los datos” [1]. El descubrimiento de patrones válidos se logra con la MD, que a través de la aplicación de diferentes técnicas puede obtener patrones y relaciones dentro de los datos permitiendo la creación de modelos. El KDD se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a estos patrones encontrados.

El proceso KDD se organiza en cinco fases: integración y recopilación de datos; selección, limpieza y transformación; minería de datos evaluación e interpretación y difusión. Si bien ambos términos, MD y KDD, se utilizan como sinónimos en realidad MD es una de las etapas del proceso de extracción de conocimiento. [2].

En esencia la MD es un mecanismo de exploración y análisis consistente en la búsqueda y extracción de información valiosa, patrones y reglas ocultas en grandes volúmenes de datos.

Las áreas de salud y epidemiología requieren de la aplicación de estrategias diferenciadas para la prevención y control sanitario. El interés principal es el análisis y estudio de la distribución geográfica de las enfermedades y su relación con los riesgos potenciales, a fin de generar planes tendientes a la prevención y control de los eventos de salud que se presentan en la población. El objetivo de los estudios epidemiológicos es determinar la relación espacio-persona-tiempo. Identificar los riesgos existentes y poder determinar las áreas afectadas es parte importante de la vigilancia epidemiológica [3].

La MDE es considerada una rama de la MD tradicional, pone énfasis en la extracción de conocimiento relevante inherente a la naturaleza espacial de los datos. La MDE provee los mecanismos y herramientas como respuesta a la

dificultad de resolver problemas de descubrimiento de conocimiento en bases de datos espaciales con los enfoques tradicionales de la minería de datos [4].

Se la puede definir como el proceso automático o semiautomático de seleccionar, explorar, modificar, visualizar y valorar grandes volúmenes de datos espaciales con el objeto de descubrir conocimiento. Permite obtener correlaciones no evidentes y potencialmente útiles entre objetos geográficos.

La aplicación de la MDE busca resolver diversos problemas mediante el descubrimiento de conocimiento, aplicando diversas técnicas donde los objetos espaciales cuentan además con características no espaciales y sirven como entrada a algoritmos de minería.

La obtención de conocimiento a partir de Bases de Datos espaciales, generadas en el ámbito de salud, permite identificar regiones donde hay que prestar especial atención en vigilancia epidemiológica y adecuar los planes de prevención y control de eventos de salud. Los avances en este sentido contribuirán a preservar y mejorar el nivel de la salud pública en general [5].

Líneas de Investigación, Desarrollo e Innovación

La línea de investigación tiene como objetivo obtener conocimiento previamente desconocido y potencialmente útil para la toma de decisiones en el área de salud tendientes a prevenir y controlar el estado de salud de la población.

A través de esta línea de investigación se propone aplicar técnicas de minería de datos espaciales para la extracción de nuevos conocimientos que asistan a la

toma de decisiones en Salud Pública en el área de influencia de la UNNOBA.

Resultados y Objetivos

Dado que es una línea de reciente creación se está llevando a cabo la etapa de recopilación de trabajos realizados en similares temáticas.

Para continuar con:

- Evaluación de métodos de KDD que tienen en cuenta las características especiales de la información de una base de datos espacial.

- Relevamiento y análisis de las herramientas existentes en la actualidad de MDE. Selección de la herramienta a utilizar en la fase de MD.

- Obtención de modelos predictivos y/o descriptivos que brinde apoyo para la toma de decisiones en el área de salud.

Formación de Recursos Humanos

En esta línea I/D se espera concluir con un trabajo de Práctica Profesional Supervisada de la Ingeniería en Informática y el desarrollo de dos tesis de posgrado dirigidas por miembros de este proyecto.

Referencias

[1] Fayyad, U.M. et al. "From Data Mining to Knowledge Discovery: An overview. Advances in Knowledge Discovery and Data Mining, Pp.1-34, AAAI/MIT Press 1996.

[2] Orallo, J.H., Quintana, M.J., Ramirez, C.. Introducción a la Minería de Datos. Pearson Prentice Hall. 2004.

[3] La minería de datos espaciales y su aplicación en los estudios de salud y epidemiología. Gonzalez Polanco, L., Perez Bentancourt, Y.G. Revista Cubana de Inf. en Ciencias de la Salud; 24(4):482-489. 2013

[4] Trends in Spatial Data Mining. Shekhar, S. Zhang, P.,Huang, Y., Vatsavai, R.. Department of Computer Science and Engineering. University of Minesota. 4-192, 200 Union ST SE, Miniapolis, MN 55455.

[5] Guasch, M.M., Piergallini,M.R., Smail, A, Russo, C., Torrigino, A., Dorzán, M.G., .Gagliardi, E. y Taranilla, M.T.. “VisualOET Software para el análisis y seguimiento de objetos espacio-temporales aplicable a Salud Pública”. CAIS (Congreso Argentino de Informática y Salud) 44 JAIIO. 2015. Pp.38-46. ISSN: 2451-7607

Análisis Comparativo de métricas de QoS de Redes Wi-Fi 802.11e con Capa Física 802.11n y 802.11ac

Santiago Pérez, Gustavo Mercado, Higinio Facchini, Alejandro Dantiacq, Gastón Cangemi
GRID TICs (Grupo UTN de Investigación y Desarrollo en TICs)
Laboratorio de Análisis de Tráfico y Seguridad
Facultad Regional Mendoza, Universidad Tecnológica Nacional
Rodríguez 273, Mendoza, Argentina
0261-5244576
[/santiagocp,gmercado,higiniofac,alejandrod,gcangemi}@frm.utn.edu.ar](mailto:{santiagocp,gmercado,higiniofac,alejandrod,gcangemi}@frm.utn.edu.ar)

Gabriel Quiroga
Sistemas Informáticos y Telecomunicaciones
Escuela de Tecnologías de la Información y las Comunicaciones
Universidad Nacional de Chilecito
9 de Julio n° 22 - Chilecito - La Rioja
0 3825-427220
hgquiroga@undec.edu.ar

Resumen

Las Redes Inalámbricas basadas en IEEE 802.11 se han vuelto las redes más populares en el acceso a los servicios WLAN corporativos y de banda ancha móvil/wireless de Internet. Para satisfacer los requerimientos de QoS (Quality of Service - Calidad de Servicio) se introdujo, a nivel de Capa 2, la tecnología IEEE EDCA (Enhanced Distributed Channel Access) 802.11e. Mientras que para satisfacer mayores velocidades de transferencias, a nivel de Capa 1, se han sucedido diversas mejoras introducidas con IEEE 802.11n y IEEE 802.11ac. Estas últimas tecnologías establecen características PHY y MAC obligatorias y opcionales.

En este proyecto de investigación se propone un análisis comparativo de las métricas de QoS de Wi-Fi entre 802.11a, 802.11n y 802.11ac, evaluando el nivel de mejoras alcanzadas con las técnicas de agregación MAC Service Data Unit (A-MSDU) y MAC Protocol Data Unit (A-MPDU), y con diferente número de streams espaciales y de ancho de banda de canal.

La experimentación se realizará mediante simulación usando una variante del modelo de estación (STA) EDCA 802.11e construido

con Redes de Petri (Universidad de Porto) combinado con el simulador Möbius (Universidad de Illinois).

Palabras claves: QoS, IEEE 802.11e, IEEE 802.11n, IEEE 802.11ac, modelación, simulación, Redes de Petri

Contexto

La línea de investigación está inserta en el proyecto homólogo, en el ámbito del Laboratorio de Análisis de Tráfico y Seguridad (Latys) del Grupo UTN GRID TICs (GRUPO UTN DE INVESTIGACION Y DESARROLLO EN TICs), del Departamento Ingeniería en Electrónica, de Facultad Regional Mendoza, de la Universidad Tecnológica Nacional. Y además, está relacionado con el trabajo de Tesis Doctoral del Mg. Ing. Santiago Pérez (de la Facultad de Ingeniería en la Universidad de Mendoza), de Maestría del Lic. Gabriel Quiroga (en la Facultad de Ciencias Exactas de la Universidad Nacional de San Juan), y de Tesis de Maestría del Ing. Higinio Facchini (de la Facultad de Informática de la Universidad Nacional de La Plata).

Introducción

En años recientes, el esfuerzo continuo en la búsqueda de comunicaciones wireless de gigabit ha dado resultados en las WLAN IEEE 802.11 [1-2]. Por ejemplo, en 2010, la Alianza Wireless Gigabit (WiGig) formada por un consorcio de líderes de la industria completó el primer borrador de la especificación WiGig [3] que definió una arquitectura unificada para habilitar comunicaciones tribanda sobre las bandas de frecuencias de 2.4, 5 y 60 Ghz.

La especificación WiGig, que apuntaba a obtener comunicaciones wireless multigigabit en la banda de 60 Ghz, contribuyó a la nueva enmienda conocida como IEEE 802.11ad. La misma se construyó sobre el estándar 802.11 existente, dando interoperabilidad con las bandas de 2.4 y 5 Ghz de las normas 802.11b/a/g/n [4] y el estándar 802.11ac [5-6].

En la banda de 5 Ghz, IEEE 802.11ac puede proveer un tasa de datos PHY máxima de cerca de los 7 Gbps. Para 80 Mhz de ancho de banda de canal, se establece un rendimiento MAC de más que 500 Mbps para un escenario con un único usuario, y un rendimiento MAC agregado de más que 1 Gbps para un escenario multiusuario.

Consecuentemente, 802.11ac puede utilizarse en los servicios de tasas de datos más altas, tales como televisión de alta definición, implementación inalámbrica de interfaces multimedia de alta definición, entre otras aplicaciones.

Este estudio pretende analizar los beneficios potenciales de QoS de 802.11ac, en términos de prestaciones MAC, comparado con 802.11a y 802.11n.

Visión general de las mejoras en las características de Capa 1 y 2 de 802.11ac

En general, 802.11ac es una extensión de 802.11n sobre las dos características básicas fundamentales conocidas: multiple input multiple output (MIMO) y ancho de banda de canal más amplio.

La idea básica es que la tasa de datos PHY máxima teórica puede crecer linealmente por un factor dado por el número de flujos espaciales (antenas de transmisión/recepción) o el ancho de banda. En otras palabras, la tasa de datos PHY puede duplicarse (cuadruplicarse) si se duplica el número de flujos espaciales o (y) el ancho de banda del canal.

A. Características PHY obligatorias y opcionales

IEEE 802.11ac mantiene la mayoría de las características PHY obligatorias de 802.11n, tales como la codificación convolucional binaria (BCC) para la corrección de errores, MIMO básico, los esquemas de modulación y codificación (MCSs) de 0 a 7, y el intervalo de guarda regular de 800 ns.

La diferencia clave en estas características obligatorias es el soporte al ancho de banda de canal de 80 MHz para lograr un incremento de prácticamente el doble de la tasa de datos de 802.11n, donde el canal más ancho tiene 40 MHz.

Como una consecuencia, 802.11ac obliga solo a un único flujo espacial, en lugar de uno o dos flujos espaciales como especifica 802.11n. Una razón para tal cambio es que el incremento del número de antenas, frecuentemente, impacta directamente en mayores costos. Por ello, los modos que utilizan más que un flujo espacial son ahora opcionales en 802.11ac. El modo de 80 MHz es una alternativa de costo inferior cuando se la compara al modo de 40 MHz con dos flujos espaciales.

En términos de características opcionales, 802.11ac define el soporte de 160 MHz de ancho de banda de canal para poder duplicar la tasa de datos sobre el ancho de banda de canal obligatorio de 80 MHz. Además, se introdujeron dos nuevos MCSs 8 y 9 basados en 256 QAM con tasas de código de 3/4 y 5/6 para una ulterior mejora de 20% y 33% en la tasa de datos, respectivamente, cuando se compara a la MCS más alta especificada en 802.11n que se basa en 64 QAM con una tasa de código de 5/6.

Con el objeto de soportar anchos de banda de canal más amplios, 802.11ac define canales de 20, 40, 80 y 160 MHz. Por ejemplo, una banda de transmisión de 40 MHz puede formarse por dos bandas contiguas de 20 MHz, mientras que una banda de transmisión de 80 MHz está formada por dos bandas contiguas de 40 MHz, en la cual una de las bandas de 20 MHz es el canal primario y el resto son canales secundarios. Sin embargo, una banda de transmisión de 160 MHz se forma por las bandas inferior y superior de 80 MHz que pueden ser contiguas o no contiguas.

B. Mejoras MAC

802.11n introdujo, como parte fundamental de su mejora MAC, dos diferentes clases de compresión por agregación de trama, llamadas A-MSDU y A-MPDU, en orden a mejorar su eficiencia MAC. También es posible combinar A-MSDU y A-MPDU, técnica conocida como agregación híbrida A-MSDU/A-MPDU [7-9].

De igual forma, la mejora MAC clave de 802.11ac está centrada alrededor de su capacidad para operaciones multicanal. En particular, 802.11ac soporta A-MSDU y A-MPDU realizada donde el tamaño de A-MSDU máximo y la longitud de la unidad de datos de servicio (PSDU) del procedimiento de convergencia PHY (PLCP) se incrementan desde los 7935 a 11426 bytes, y de 65535 a 1048576 bytes, respectivamente. Esto implica una ulterior mejora de su eficiencia MAC en base a tasas de datos PHY más altas.

Además, sumado a su capacidad multicanal, 802.11ac soporta protección realizada en la que el mecanismo handshake RTS/CTS ha sido modificado para soportar reserva de ancho de banda estática y dinámica, y transporta información de ancho de banda de canal. La idea es que las tramas RTS y CTS se transmiten por la STA VHT (Very High Throughput) sin usar alto rendimiento (non-HT) duplicando la unidad de datos del protocolo PLCP (PPDU). La duplicación de una transmisión no HT de 20 MHz en cada canal adyacente de 20 MHz de un ancho de banda de canal más amplio suministra

compatibilidad hacia atrás con otros dispositivos 802.11. De esta forma, las STAs heredadas pueden decodificar las tramas RTS y CTS, y actualizar su vector de asignación de red (NAV) para prevenir el problema de terminales ocultas que escalará con las operaciones multicanal de 802.11ac, en los canales secundarios de una STA VHT. Adicionalmente, la información del ancho de banda del canal en las tramas RTS y CTS junto con la canalización 802.11ac habilita la vecindad entre STAs VHT para ganar conocimiento de los canales secundarios de las STAs VHT.

El 802.11ac también soporta capacidades VHT, tales como: múltiple entrada múltiple salida multiusuario descendente (MU-MIMO) y el ahorro de potencia de las oportunidades de transmisión (TXOP), a través de su campo de información de capacidades VHT de la trama de administración. Las transmisiones MU-MIMO descendente pueden organizarse en la forma de MU-TXOP para facilitar el compartimiento de TXOP, donde el AP puede realizar transmisiones simultáneas a múltiples STAs receptoras usando su ID de grupo. El MU-TXOP requiere al menos una única STA, indiferente de su categoría de acceso (AC), para recibir tráfico desde la AC que ha obtenido exitosamente la TXOP, en lugar del requerimiento EDCA original donde sólo se permite una transmisión originada por la AC que gana la TXOP.

Finalmente, el ahorro de potencia TXOP se introdujo para que la STA VHT no AP puede entrar en el modo de ahorro de energía, cuando confirma que no es el destino durante de esa TXOP.

Experiencias de Simulación

Para entender los beneficios de 802.11ac se realizarán ensayos experimentales por simulación que servirán para comparar su performance MAC en términos de rendimiento máximo. Al efecto se proponen los siguientes escenarios:

- Canal de 40 MHz con 2×2 MIMO
- Canal de 80 MHz con single input, single output (SISO)
- Canal de 80 MHz con 2×2 MIMO
- Canal de 160 MHz con SISO

Con dicho objeto se tienen en cuenta las siguientes suposiciones:

- Transmisión punto a punto de un transmisor y un receptor operando con el modo EDCA (simple AC).
- Cada trama tiene un tamaño fijo.

La evaluación experimental de WiFi EDCA 802.11e, IEEE 802.11n y IEEE 802.11ac se utilizará un modelo de simulación implementado en Redes de Actividades Estocásticas Jerárquicas (HSAN), que se ejecuta sobre el simulador Möbius [10]. Las HSAN son una versión de las Redes de Petri. Este modelo comprende una implementación precisa y detallada de la función EDCA asociada a las estaciones con QoS, considerando tanto su perspectiva funcional y temporal. Además, y desde el punto de vista de la modelación pueden obtenerse un gran número de medidas de métricas de QoS. El modelo puede ser usado como una estructura base para construir modelos más complejos y de más alto nivel.

Esto provee una importante flexibilidad en el proceso de evaluación, como la aceleración en el análisis de diferentes escenarios de red. Todas las simulaciones experimentales se obtendrán usando el modelo de simulación EDCA, previamente descrito, con un intervalo de confianza del 95% y una precisión del 5 %.

Las métricas de QoS a analizar son: rendimiento absoluto, rendimiento relativo, pérdida de paquetes, retardo de cola promedio y tamaño de cola promedio.

Experiencias de Laboratorio

Para el contraste se realizarán experimentos de laboratorios pilotos a implementar para el proyecto con el siguiente equipamiento, de los Labs del GRID TICs:

- 1 (un) AIRPCAP NX: USB 802.11 a/b/g/n Adapter (capture + injection)
- 1(un) AP 802.11ac
- 2 (dos) Placas USB Wireless Linksys WUSB600N 2 (dos) Placas USB Wireless Kozumi K300MWUN
- 2 (dos) Placas Routerboard Mikrotik RB433AH
- 2 (dos) Placas MiniPCI Wireless Mikrotik R52N
- 2 (dos) Gabinetes para RB433
- 2 (dos) Antenas para Indoor Mikrotik
- 1 (un) Wireless Access Point Sp918gk Micronet
- 1 (un) Switch KVM 4 port Trendnet TK-400K
- 1 (un) generador de Tráfico IP Traffic

Objetivos

A. Objetivo general:

Determinar cuantitativamente el impacto en las métricas de QoS, de las Redes Wi-Fi IEEE 802.11e, las mejoras introducidas en la norma IEEE 802.11ac respecto a la norma IEEE 802.11n, en el punto de rendimiento máximo.

B. Objetivos específicos:

- a) Determinar los escenarios de prueba o experimentación para el análisis comparativo.

b) Establecer la modalidad de comparación de los umbrales de las métricas de QoS y para cada uno de los escenarios de experimentación en el punto de rendimiento máximo.

c) Determinar las configuraciones de los protocolos IEEE 802.11e, IEEE 802.11n y 802.11ac para cada uno de los escenarios de experimentación, y

Líneas de investigación y desarrollo

El presente proyecto de investigación implica el afianzamiento de esta línea de estudio, iniciada en diversos proyectos internos de GRID TICs.

Los integrantes del Proyecto, son miembros del Grupo UTN GRID TICs, y han trabajado en distintos proyectos en el área de las TICs, desde hace 15 años en las temáticas de análisis de tráfico, redes avanzadas, Internet Version 6, modelación de tráfico con Redes de Petri, Redes de Sensores Inalámbricos, etc. El presente proyecto articula y complementa el aspecto de QoS (Calidad de servicios) en redes Wi-Fi, del PID de GRID TICs, dirigido por el Ing. Santiago Pérez llamado “MEDIDA, MODELACIÓN Y SIMULACIÓN DEL TRÁFICO DE VIDEO EN REDES WI-FI 802.11E - ALGORITMO MTDA”, Código UTN-Incentivos: UTI3502TC, Disposición de la. Secretaría de Ciencia y Tecnología UTN 570/14, Fecha de Inicio: Ene 2015, Fecha de Finalización: Dic 2017.

Se aporta el conocimiento y experiencia de algunos de sus miembros en el tema de Redes WLAN 802.11, debido a su condiciones de Magister y Especialistas en Redes de Datos de la Universidad Nacional de La Plata, y por ser instructores de los cursos oficiales avanzados de Redes de Cisco de la carrera de CCNP y de CCNA (específicos de las temáticas de redes en todos los niveles del modelo OSI).

Formación de Recursos Humanos

El equipo de trabajo está integrado por docentes investigadores, y becarios graduados y alumnos de la Universidad Tecnológica

Nacional, Facultad Regional Mendoza y de la Universidad Nacional de Chilecito. El Mg. Ing. Santiago Perez ha desarrollado y presentado para defensa la tesis doctoral sobre la modelación de las redes wireless 802.11e. Mientras que el Ing. Higinio Facchini y el Lic. Gabriel Quiroga, desarrollan actualmente sus tesis de Maestría sobre tráfico de video en redes Wi-Fi IEEE 802.11e.

Referencias

- [1] IEEE Std 802.11™-2012
- [2] G. Hiertz, D. Denteneer, L. Stibor, Y. Zang, X.P. Costa, and B. Walke, “The IEEE 802.11 universe”, IEEE Communications Magazine, 48(1):62–70, January 2010.
- [3] Wireless Gigabit Alliance, “Defining the future of multi-gigabit wireless communications”, White Paper, pages 1–5, 2010.
- [4] IEEE 802.11n 2009. Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, amendment 5: Enhancements for higher throughput. October 2009.
- [5] IEEE P802.11ac/D5.0, January 2013
- [6] E. H. Ong, et. al., “IEEE 802.11ac: Enhancements for Very High Throughput WLANs”, IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2011
- [7] E. Perahia and R. Stacey, “Next Generation Wireless LANs: Throughput, Robustness, and Reliability in 802.11n”, Cambridge University Press, The Edinburg Building, Cambridge, UK, 2008.
- [8] D. Skordoulis, Q. Ni, H. Chen, A. P. Stephens, C. Liu, and A. Jamalipour, “IEEE 802.11n MAC frame aggregation mechanisms for next-generation high-throughput WLANs”, IEEE Wireless Communications, 15(1):40–47, February 2008.
- [9] J. Kolap, S. Krishnan, N. Shaha, “Frame Aggregation Mechanism For High-Throughput 802.11N WLANs”, International Journal of Wireless & Mobile Networks (IJWMN) Vol. 4, No. 3, June 2012
- [10] <https://www.mobius.illinois.edu>

Evaluación del Enrutamiento y Acceso al Medio en Redes Mesh Auto-Configurables

Daniel Britos¹, Laura Vargas¹, Silvia Arias¹, Fernando Menzaque², Sergio Chalave¹, Nicolás Echániz³, Antonella Sgarlatta¹, Agustín Moreno¹

dbritos@efn.uncor.edu, laura.monica.vargas@unc.edu.ar, edith_edit@gmail.com, menzaque@gmail.com, chalave@gmail.com, nicoechaniz@altermundi.net, antosgar@gmail.com, agumoreno12@gmail.com

¹Laboratorio de Redes y Comunicaciones de Datos, Departamento de Computación, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba

²Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba

³Asociación Civil AlterMundi, José de la Quintana, Córdoba

I. RESUMEN

Las redes comunitarias de tipo Mesh mediante la utilización de la tecnología inalámbrica (802.11b/g/n) permiten el acceso libre a la red a través de la colaboración de los propietarios que dejan abiertos sus puntos de acceso a cada nodo. Son particularmente útiles en aquellos lugares que no cuentan con una infraestructura de comunicaciones comercial. En este proyecto se pretende estudiar el problema de escalabilidad en estas redes, utilizando simuladores especialmente diseñados para ello. Se ahondará en el problema de contención en el acceso al medio que degrada la performance de estas redes brindando soluciones a través de diversos protocolos.

Palabras clave: redes mesh, redes libres, redes inalámbricas, protocolos de encaminamiento, redes comunitarias.

II. CONTEXTO

El presente trabajo se realiza en el ámbito del Laboratorio de Redes y Comunicaciones de Datos que funciona en el Departamento de Computación perteneciente a la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de Córdoba. Este trabajo es continuación del presentado en el período 2014-2015 ante la SECyT (Secretaría de Ciencia y Técnica) de la UNC (Universidad Nacional de Córdoba). Esta investigación previa ha dado lugar a varias publicaciones y presentaciones en distintos congresos tales como CACIC 2015 con el trabajo “BATMAN Adv. Mesh Network Emulator”, publicado en el Libro de Actas y WICC 2014

“Integración de Redes Comunitarias con Redes Universitarias”, entre otros.

La Asociación Civil AlterMundi participante de este proyecto ha obtenido el año pasado el premio FRIDA ofrecido por LACNIC (Latin America and Caribbean Network Information Centre) en la categoría “Dispositivos, Infraestructura y Tecnologías. Aceleración y Expansión del Acceso” (ver <http://programafrida.net/awards>).

III. INTRODUCCIÓN

De acuerdo a lo propuesto en el proyecto SECyT- UNC 2014-2015 Código 05/M246, “Estudio de la escalabilidad del enrutamiento híbrido (L2 + L3) en redes mesh autoconfigurables”, se desarrolló durante estos dos últimos años un simulador de redes mesh que permitió estudiar distintas topologías y analizar su escalabilidad. Para topologías de muchos nodos, los requerimientos computacionales necesarios son grandes y no se encuentran disponibles en las pequeñas comunidades donde se despliegan estas redes por lo que se hicieron pruebas de acceso remoto, pero debido al ancho de banda de las conexiones esto no resultó práctico. Además la interfaz de usuario del simulador no reflejaba la distribución geográfica de los nodos. Lo expuesto generó los requerimientos para una nueva versión del simulador, el cual debería ser accesible a través de la web en la cual los nodos deberían estar geolocalizados sobre un mapa. Por otro lado, se detectó que el comportamiento de la red en el simulador era mejor que la red real desplegada en Quintana, se analizaron las causas y se llegó a la conclusión de que el simulador no

tenía en cuenta la contención del acceso al medio que provoca un gran número de colisiones en ciertos lugares de la red [1] [2]. Para esta nueva etapa del proyecto se agrega el estudio de esta problemática.

Internet a pesar de su difusión global, no cuenta con acceso igualitario en todo el planeta. Esta brecha de acceso a la información, generalmente perjudica a sectores empobrecidos, y/o a poblaciones geográficamente inaccesibles. En Córdoba, esta situación afecta a regiones serranas, donde es difícil obtener señal telefónica o bien esta es inexistente. En el oeste cordobés, en la región de las Sierras Grandes, este paradigma afecta a la comunidad educativa de escuelas como la de Pampa de Achala, los Gigantes, San Jerónimo, Sagrada Familia, Tala Cañada y la Sierrita.

En los últimos años, el Laboratorio de Redes y Comunicaciones de Datos de la FCEFyN de la UNC (LaRyC) ha estudiado este problema de aislamiento, concluyendo que la mejor opción para darle solución, son las redes libres del tipo mesh.

Luego de conversaciones e intercambios con grupos similares pertenecientes a la Universidad Politécnica de Cataluña (UPC) y a la Universidad Politécnica de Berlín (TUB), se concluyó que el difícil acceso a los posibles nodos requiere mejorar algunos aspectos de las redes mesh tales como: la auto-configuración de los nodos, la contención del acceso al medio y la escalabilidad de la red.

Las redes inalámbricas comunitarias que cuentan con miles de nodos, han sido hasta ahora dependientes de la implementación del modo infraestructura y por lo general con uso de software propietario en sus routers.

Las grandes redes como guifi.net [3], con 20.000 nodos activos o AWMN [4], con 2.400 nodos activos, hacen un uso intensivo de hardware Ubiquiti y Mikrotik, corriendo los sistemas operativos propietarios AirOS [5] y RouterOS [6], respectivamente.

Las redes mesh *ad-hoc* auto-configurables basadas en software de código abierto nunca han llegado a tal escala. Proyectos exitosos como Village Telco utilizan un modelo mixto en el que la malla *ad-hoc* desplegada a “nivel de calle”

utiliza software libre, pero los enlaces troncales utilizan el modo infraestructura.

Estas nubes a “nivel de calle”, que utilizan B.A.T.M.A.N. Avanzado (Nivel 2) para el enrutamiento dinámico, cuentan por lo general con menos de un centenar de nodos. La columna vertebral de la infraestructura utiliza hardware dedicado, ejecutando diferentes programas en la nube de routers, por lo general de software propietario.

Las redes comunitarias de todo el mundo buscan una mayor participación de los usuarios, para facilitar la autogestión, minimizando o eliminando la necesidad de configuración manual. Sin embargo, las herramientas corrientes de configuración automática son generalmente útiles sólo en pequeñas redes. Estas herramientas no pueden hacer frente a las redes que crecen o comienzan a conectarse a las “nubes” vecinas.

Las redes mesh *ad-hoc* en la comunidad inalámbrica, tradicionalmente se han basado en firmware de redes mesh auto-configurables para routers de un solo radio. Existen numerosos ejemplos de estas: SECN de Village Telco [7], Nightwing [8], Robin [9], por nombrar algunos. Nodos mesh de un solo radio se enfrentan a una degradación de rendimiento grave para trayectos de múltiples saltos, lo que limita en gran medida la escalabilidad de la malla [10].

Los nuevos firmware de malla de cero-configuración diseñados para nodos multi-radio han superado con éxito esta cuestión. Un ejemplo de ello es AlterMesh [11], desarrollado por Altermundi. El enrutamiento dinámico del protocolo usado en AlterMesh es BATMAN Avanzado (batman-adv) [12] que implementa la interfaz alterna, modo en el cual cuando un paquete llega a través de una interfaz elige otra interfaz diferente de salida, si esta existe, con una calidad de enlace comparable.

Existen redes en Argentina desplegadas con este modelo, el uso de routers off-the-shelf y el firmware AlterMesh, son casos de éxito de las redes comunitarias de radio múltiple y de cero configuración. Mientras que esta solución representa un avance importante y ha permitido la expansión de las redes comunitarias en América Latina, todavía tiene problemas de escalabilidad que son imposibles de superar con el modelo existente. El principal problema radica en la

interconexión de las diferentes nubes de capa dos (L2). Las redes batman-adv pueden manejar varias puertas de enlace predeterminadas, pero el protocolo no está diseñado para manejar diferentes rutas de acceso a otras nubes L2 de manera óptima.

En la primera etapa se intentó representar las redes mesh en diferentes escenarios de pruebas con el simulador de redes GNS3, pero se determinó que no existía una forma eficiente de simular los enlaces inalámbricos, uno de los principales objetivos perseguidos, por este motivo se descartó esta opción. Se procedió entonces a estudiar las posibilidades de utilizar el simulador de redes NS3, para lograr esto se empezó a desarrollar un simulador basado en el NS3 y como primera medida se intentó implementar el módulo Batman Advanced, pero después de dos meses de trabajo no se logró integrar el módulo al NS3. Existía aún la posibilidad de desarrollar el módulo Batman Advanced en modo kernel para el NS3 pero como se perdía la facilidad de simularlo cada vez que salía una nueva versión, se optó por desarrollar el simulador desde cero. El simulador desarrollado se basó en VirtualBox, y en Virtual Switch, con esta última herramienta se logró superar el escollo de la emulación de los enlaces inalámbricos.

Además del desarrollo informático necesario para mejorar esta tecnología, la autogestión de las redes comunitarias requiere la capacitación de los usuarios y el crecimiento de estas redes también está sujeto a su conocimiento. Con el fin de apoyar la difusión de éste tipo de prácticas, el proceso se documentará mediante producciones audiovisuales y se generarán spots para su publicación masiva. Asimismo, se producirán videos tutoriales para colaborar en la capacitación de los usuarios.

IV. OBJETIVOS Y LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los objetivos de este proyecto de investigación los podemos dividir en:

A. Objetivo General

Implementar redes mesh en el noreste cordobés para favorecer la accesibilidad a medios de comunicación. Una de las principales dificultades de este objetivo es que las

comunidades del oeste cordobés no cuentan con suministro de energía eléctrica.

B. Objetivos Específicos

Mejorar la tecnología de redes mesh para que pueda funcionar con la mínima atención técnica posible.

Estudiar la posibilidad de mejorar el modo de contención en acceso al medio.

Evaluar el potencial del ruteo híbrido en capa 2 y capa 3 del modelo OSI de redes mesh.

Aprovechar las capacidades de auto-configuración del firmware libre-mesh.

Difundir los resultados para realimentar el proceso de desarrollo de los protocolos libre-mesh.

Formación de recursos humanos específicos en las áreas objeto de estudio.

MATERIALES Y MÉTODOS

Materiales que se utilizarán en las Experiencias

Los materiales que se utilizarán en las experiencias son: computadoras y súper computadoras para realizar las simulaciones, nodos wireless para realizar el banco de prueba y luego nodos completos con antenas paneles solares, baterías y torres para desplegar en el oeste Cordobés. Software libre C, GNS3 y NS3.

Métodos y Técnicas a Utilizar

Este estudio se estructurará como una colección de experimentos que ayudarán a medir una serie de métricas en un determinado conjunto de escenarios utilizando diferentes modelos de red.

Los protocolos de enrutamiento de todos los nodos participantes y toda la comunicación dentro de una nube L2 será manejado por el protocolo batman-adv mientras que la comunicación, con destino u origen fuera de la nube L2, se enrutará por el protocolo bmx6 [13] Batman Experimental versión 6 [14].

Se implementarán las siguientes alternativas:

-La "Arquitectura A" que se define como una red en donde ambos protocolos se ejecutan en cada nodo.

-La "Arquitectura B" que se define como una red donde el protocolo "bmx6" solamente se ejecuta en los nodos frontera los cuales están

conectados a otras nubes o poseen Uplinks a Internet.

El “Demonio de encaminamiento de presencia de Capa 3” en este estudio representa una progresión de la arquitectura A a la arquitectura B.

El objetivo principal de experimentar estas alternativas es predecir la escalabilidad de cada solución, mientras se proporciona una visión sobre su desempeño y deficiencias.

El rendimiento, la escalabilidad y la ruta serán los indicadores de selección y de convergencia cuando se ejecuten en estos escenarios los diferentes modelos de redes.

Se estudiará la convergencia con el software desarrollado en la primera etapa del proyecto, en este software es difícil realizar métricas de rendimiento por lo tanto se deberá buscar la forma de realizar esto ya sea mejorando el simulador ya desarrollado o una de las alternativas es medir el rendimiento con el software NS3.

Se elaborarán módulos de Batman Advanced para NS3, donde se simulará el ambiente de trabajo y el comportamiento del protocolo a escalas del orden de 10 y 100 nodos.

Una vez completadas las pruebas en el simulador se realizarán en la red del laboratorio con 10 nodos. Luego se pasará a probarlas en las redes de la Quintana y Wibed [15]. Las pruebas en los distintos escenarios implicarán ajustes dinámicos y estáticos. Las configuraciones dinámicas serán creadas a través de los nodos mediante simulaciones de fallos, así como con nodos móviles cuando sea posible.

VI. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está formado por cinco investigadores categorizados por el Programa Nacional de Incentivos (categorías III y IV), de los cuales uno es Doctor en Ciencias de la Ingeniería, otro Doctor en Matemática, un tercero Magister en Ciencias de la Ingeniería mención Redes, y dos son Especialistas en Docencia Universitaria. También integran el equipo dos alumnos que realizarán su tesis final de grado en carreras de Ingeniería. Además uno de sus miembros dirigirá el Doctorado en Ciencias de la Ingeniería de un Ingeniero en Comunicaciones egresado de la Universidad Blas Pascal, Magister

en Comunicaciones, egresado del Instituto Politécnico de Torino.

Esta línea de trabajo desarrollada en el Laboratorio de Redes y Comunicaciones de Datos permitirá la capacitación en el tema de los alumnos de las carreras de Ingeniería en Computación e Ingeniería Electrónica. Esto tendrá un efecto multiplicador produciendo egresados más capacitados que brinden mejores servicios a la industria y a la sociedad en un área que como la de Redes Inalámbricas está en creciente desarrollo.

REFERENCIAS

- [1] N. Nandiraju, D. Nandiraju, L. Santhanam, B. He, J. Wang, and D. Agrawal, “Current challenges and future directions of web-in-the-sky”, in *Wireless Communications, IEEE*, vol. 14, no. 4, pp. 79–89, August 2007.
- [2] A. Acharya, A. Misra, and S. Bansal, “Design and analysis of a cooperative medium access scheme for wireless mesh networks” in *Broadband Networks, 2004. BroadNets 2004. Proceedings. First International Conference on*, pp. 621–631, Oct 2004.
- [3] “guifi.net - xarxa de telecomunicacions oberta, lliure i neutral guifi.net” [Online]. Available: <http://www.guifi.net/>
- [4] “AWMN WiND - wireless nodes database” [Online]. Available: <http://wind.awmn.net/>
- [5] “AirOS v | ubiquiti networks, inc.” [Online]. Available: <http://www.ubnt.com/airos>
- [6] “MikroTik routers and wireless” [Online]. Available: <http://www.mikrotik.com/software.html>
- [7] “Village telco” [Online]. Available: <http://villagetelco.org/>
- [8] “Nightwing by: Lugro-mesh” [Online]. Available: <http://nightwing.lugro-mesh.org/ar/>
- [9] “ROBIN - open source mesh network: Index” [Online]. Available: <http://robin.forumup.it/index.php?mforum=robin>
- [10] J. D. Britos, “Multiple hop wireless link” *Latin America Transactions, IEEE*, vol. 3, no. 4, pp. 1–7, 2005.

- [11]“HomePage - altermundi.” [Online]. Available: <http://http://www.altermundi.net/>
- [12]“Doc-overview - batman-adv - open mesh” [Online]. Available: <http://www.open-mesh.org/projects/batman-adv/wiki>
- [13]“Doc-overview - bmx6” [Online]. Available: <http://http://bmx6.net/projects/bmx6>

- [14]“HomePage - wireless battle of the mesh” [Online]. Available: <http://battlemesh.org/>
- [15]P. ESCRICH, R. BAIG, A. NEUMANN, A. FONSECA, L. NAVARRO, and F. FREITAG, “Wibed, a platform for commodity wireless testbeds” Nov 2013.

EFICIENCIA ENERGETICA EN DISPOSITIVOS MÓVILES PARA FACILITAR SU USO EN ZONAS RURALES AISLADAS

Sergio Rocabado¹, Ernesto Sanchez¹, Susana Herrera², Carlos Cadena³

(1) *Centro de Investigación y Desarrollo en Informática Aplicada (CIDIA)*
Facultad de Ciencias Exactas, Universidad Nacional de Salta
srocabad@cidia.unsa.edu.ar, esanchez@cidia.unsa.edu.ar

(2) *Instituto de Investigación en Informática y Sistemas de Información (IISI)*,
Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero
sherrera@unse.edu.ar

(3) *Instituto de Investigaciones en Energía No Convencional (INENCO)*,
Universidad Nacional de Salta y Consejo Nacional de Investigaciones Científicas y Técnicas
cadenacinenco@gmail.com

Resumen

Las zonas rurales aisladas del país se caracterizan, entre otros aspectos, por su baja densidad demográfica, población dispersa, cobertura de red celular muy limitada y carencia de servicio de distribución de energía eléctrica. Los habitantes de estas zonas utilizan energías alternativas, como paneles solares y grupos electrógenos, para cubrir necesidades energéticas elementales. La región Noroeste de Argentina (NOA) posee numerosas zonas de este tipo, donde las posibilidades de acceso a la información digital son prácticamente nulas debido a que es muy difícil suministrar energía eléctrica a los equipos computacionales.

Por su bajo consumo energético, respecto de computadoras convencionales, los dispositivos móviles (celulares, smartphones, tablets y otros) se constituyen en una alternativa viable para este tipo de zonas.

El grupo de trabajo lleva adelante una investigación aplicada que busca reducir el consumo de energía de los dispositivos móviles para facilitar su utilización en zonas rurales aisladas abastecidas con energía solar fotovoltaica, posibilitando el acceso a información digital a pobladores de estas zonas sin necesidad de que tengan que trasladarse a centros urbanos.

La propuesta resultante tiene posibilidad real y concreta de ser replicada en otras zonas rurales del país que tengan características similares.

Palabras clave: Dispositivos móviles, redes móviles, consumo de energía, zonas rurales aisladas, energía solar fotovoltaica.

1 Contexto

El presente trabajo se lleva a cabo en el marco del proyecto de investigación N° 2268/0 “Despliegue seguro de MANETs en zonas rurales de recursos energéticos limitados” en colaboración con el proyecto de investigación “Energías solar fotovoltaica y eólica: desarrollo y transferencia de equipos a pobladores de zonas rurales de la Provincia de Salta, y su impacto en la calidad de vida”, ambos financiados por el Consejo de Investigación de la Universidad Nacional de Salta.

El equipo de investigación se encuentra conformado por investigadores de diferentes institutos del NOA: Instituto de Investigación en Informática y Sistemas de Información (IISI) de la Facultad de Ciencias Exactas de la UNSE, Instituto de investigaciones en energía no convencional (INENCO) y Centro de Investigación y Desarrollo en Informática Aplicada (CIDIA) de la UNSa. Además, se cuenta con la colaboración y asesoramiento del Laboratorio de Investigación en Nuevas Tecnologías Informáticas (LINTI) de la UNLP, por intermedio del Lic. Javier Díaz, director de tesis de doctorado de uno de los integrantes del proyecto.

2 Introducción

El recurso energético natural de la región del NOA (Noroeste Argentino) es la energía solar. La aridez del clima y la latitud tropical hacen que se cuente con una alta radiación solar la mayoría de los días del año. Éste es un recurso limpio y renovable que puede ser aprovechado en poblaciones rurales aisladas del NOA que se encuentran fuera del alcance de los centros de distribución de energía [1].

Si bien los sistemas fotovoltaicos actualmente son utilizados en muchas zonas aisladas, generalmente en comunidades rurales donde tuvo intervención el PERMER [2]; el aprovechamiento de la energía solar

para la carga de dispositivos móviles no está lo suficientemente difundida. En el desarrollo de este trabajo se introduce la aplicación de esta tecnología, contribuyendo a mejorar las posibilidades de comunicación de los pobladores de la región.

El trabajo se encuentra dividido en 4 etapas: 1) Determinar los requerimientos de energía de un celular para ejecutar tareas utilizadas en zonas rurales, 2) Verificar que el sistema fotovoltaico se ajuste a los requerimientos energéticos del celular, 3) Comunicar el celular con el sistema fotovoltaico para optimizar la carga en función de los requerimientos del dispositivo y 4) Optimización de la potencia entregada por el sistema fotovoltaico.

Etapla 1: Determinar los requerimientos de energía de un celular para ejecutar tareas utilizadas en zonas rurales.

A continuación se enumeran los pasos para establecer el consumo energético del dispositivo móvil:

1. Seleccionar tareas de uso frecuente en zonas rurales del NOA, entre otras se pueden mencionar: Navegación Web, acceso a correo electrónico, Geo-Posicionamiento, Reproducción de audio y video.
2. Asignar aplicaciones móviles para ejecutar las tareas, algunos ejemplos: Navegación Web (Google Chrome, Mozilla firefox, Opera Mini), Mensajería instantánea (Whatsapp, Facebook Messenger, Google Hangout), GPS Tracking (Geo tracker, GPS hiking, My track).
3. Determinar el consumo energético de las aplicaciones para realizar la tarea asignada, el consumo se mide utilizando alguno de los siguientes mecanismos: a) Sistemas que miden la corriente eléctrica que circula hacia el dispositivo, utilizando hardware externo (multímetro y batería abierta) y software para automatizar las mediciones [3], b) Estimaciones utilizando un modelo de potencia [4], c) Conexiones a nivel circuito [5] y d) Sensores internos del dispositivo [6].
4. Seleccionar la aplicación que consuma menor energía para ejecutar la tarea asignada (eficiencia energética).
5. Agrupar las tareas en perfiles, con base en los requerimientos informáticos de los pobladores y/o visitantes de la zona.
6. Determinar el consumo energético de cada perfil, en función del consumo de energía de las aplicaciones elegidas para ejecutar las tareas del perfil.

El requerimiento de energía del celular dependerá del perfil utilizado en la zona rural.

Etapla 2: Verificar que el sistema fotovoltaico se ajuste a los requerimientos energéticos del celular

Se calcula la energía que el sistema fotovoltaico es capaz de entregar al dispositivo móvil, en función de: factores climáticos de la zona de despliegue (radiación solar y

temperatura), características técnicas del panel y un factor de riesgo para condiciones climáticas desfavorables.

Conociendo la potencia suministrada por el sistema fotovoltaico, se establecen cuales son los perfiles que se podrán ejecutar cuando el celular se encuentre en la zona rural aislada.

A continuación se ilustran tres situaciones que pueden ocurrir con frecuencia:

En la Fig. 1 se observa el consumo de energía del celular (C_0) cuando se encuentra conectado pero en reposo a la espera de recibir órdenes (“stand by”), el remanente de energía (R) se puede aprovechar para ejecutar algún perfil de aplicaciones, en función de los requerimientos del usuario.

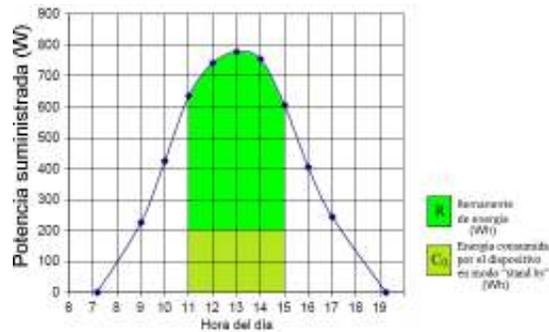


Figura 1 - Celular sin ejecutar aplicaciones, modo “stand by”

En la Fig. 2 se ilustra el consumo de energía del celular cuando se ejecuta un perfil de aplicaciones (P_1), en el gráfico se observa que el sistema fotovoltaico entrega energía suficiente para ejecutar el perfil 1, de 11 a 15 (4 horas). El remanente (R) puede ser utilizado para cargar la batería del dispositivo móvil.

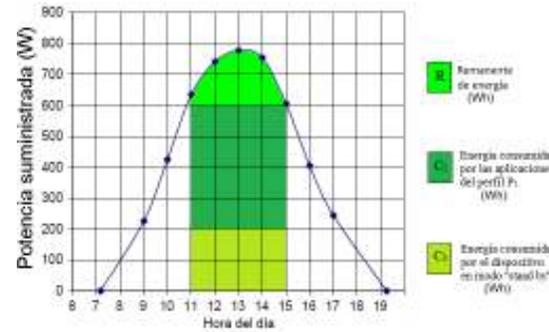


Figura 2 - Celular ejecutando aplicaciones sin déficit de energía

En la Fig. 3 se presenta un ejemplo de ejecución de perfil durante 4 horas (14 a 18), se observa un déficit de energía a partir de horas 15, déficit que se puede evitar utilizando la energía almacenada en la batería del celular. En condiciones climáticas favorables se puede estimar el tiempo de reposición del “préstamo energético”.

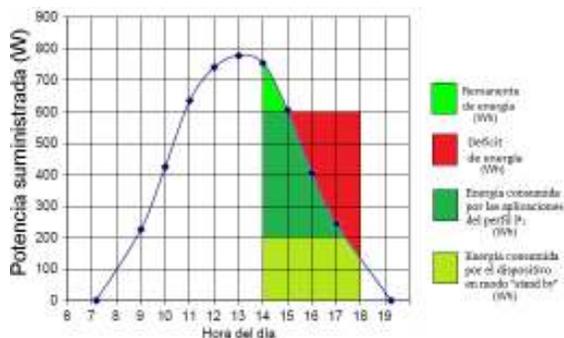


Figura 3 - Celular ejecutando aplicaciones con déficit de energía

Etapa 3: Comunicación entre el dispositivo móvil y el sistema fotovoltaico.

En esta etapa se tiene previsto agregar al cargador portátil una placa Arduino UNO [7] y sensores para medir la radiación actual y la temperatura de los módulos fotovoltaicos, los valores de radiación y temperatura serán recolectados periódicamente por la placa Arduino y almacenados en una memoria flash.

Cuando el dispositivo móvil se encuentre en zona rural aislada y requiera energía para ejecutar aplicaciones, se comunicará con el modulo Arduino del panel solar a través de la interfaz de radio Bluetooth [8] y solicitará información de la capacidad del sistema fotovoltaico.

La placa Arduino, en función de parámetros climáticos actuales e información histórica acumulada en la memoria flash, estimará la cantidad de energía que el cargador solar puede suministrar en las siguientes horas del día y enviará este valor al celular.

La información entregada al celular permitirá determinar:

- El tiempo de carga en modo “stand by”.
- El tiempo de carga de la batería cuando se ejecuta un perfil de aplicaciones con remanente de energía.
- Los perfiles de aplicaciones “sin” déficit de energía que se pueden ejecutar y por cuanto tiempo.
- Los perfiles de aplicaciones “con” déficit de energía que se pueden ejecutar utilizando la carga almacenada en la batería del celular y durante cuanto tiempo.
- El tiempo que se requiere para reponer la energía en la batería del dispositivo, cuando se ejecuta un perfil que genera déficit de energía.

Además, se pueden activar mecanismos de ahorro de energía en el celular cuando el cargador se encuentre en situaciones desfavorables para la adquisición de energía.

Etapa 4: Optimización de la potencia entregada por el sistema fotovoltaico

El voltaje que suministra un sistema fotovoltaico, generalmente, es superior al voltaje requerido por la batería del celular, por esta razón es necesario utilizar un

circuito regulador de voltaje. Se propone la modificación del sistema fotovoltaico conectando celdas en paralelo, lo que genera un incremento de la corriente y la disminución del voltaje a la salida del panel, y en consecuencia se consigue un incremento de la potencia entregada al celular a la salida del regulador (acelerando la carga).

Para ilustrar esto, en la Fig. 4 se presentan dos sistemas fotovoltaicos de 4 celdas, donde cada celda es capaz de suministrar 5 V y 1 A. La carga de una batería estándar de celular (3,7 V) se optimiza utilizando la configuración en paralelo, ya que entrega un voltaje ligeramente superior al requerido (5V) y un valor elevado para la corriente (4A).

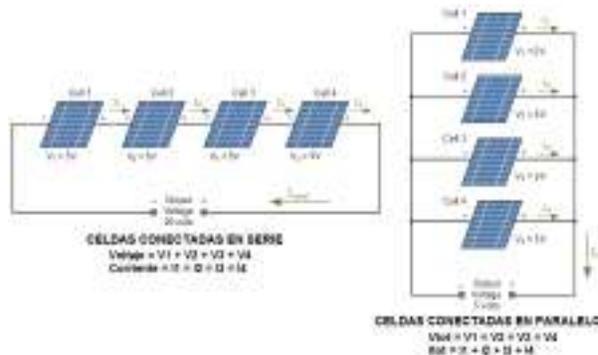


Figura 4 - Configuración de celdas fotovoltaicas en serie y paralelo

3 Líneas de Investigación, Desarrollo e Innovación

Las principales líneas de investigación de esta propuesta son:

- Computación móvil. Dispositivos móviles. Arquitectura y Sistemas Operativos.
- Despliegue de redes móviles en zonas aisladas.
- Eficiencia energética en dispositivos móviles. Técnicas para administrar y conservar energía.
- Modelos de consumo de energía para dispositivos móviles.
- Mecanismos para medir el consumo detallado de energía (Fine-Grained) en dispositivos móviles. Medición por software. Medición utilizando hardware externo.
- Baterías para dispositivos móviles. Proceso de carga. Tiempo de carga. Duración. Cargadores.
- Energía solar fotovoltaica.
- Sistemas de adquisición de energía solar fotovoltaica.
- Cargadores fotovoltaicos portátiles para dispositivos móviles.

4 Objetivos y Resultados

En este trabajo se estudia una solución tecnológica como alternativa para satisfacer una necesidad social, por esta razón los objetivos generales de esta investigación se plantean desde dos perspectivas diferentes, una tecnológica y otra social.

Objetivo Tecnológico:

- Realizar una contribución para el uso de dispositivos móviles en zonas remotas con recursos energéticos limitados.

Objetivo social:

- Posibilitar el acceso a la información y al conocimiento a pobladores de zonas rurales aisladas que se encuentran fuera del alcance de los centros de distribución de energía.

Para ello, se persiguen los siguientes objetivos específicos:

- Diseñar estrategias para el despliegue de redes móviles en zonas remotas, haciendo uso de energía solar fotovoltaica para la recarga de dispositivos móviles y de comunicación.
- Promover el uso de dispositivos móviles y de comunicación en comunidades rurales aisladas abastecidas con energía solar fotovoltaica.
- Elaborar una propuesta para reducir el consumo energético en dispositivos móviles sin afectar su rendimiento y confiabilidad.
- Construir un modelo de consumo energía que permita determinar en forma detallada (Fine-grained) la potencia utilizada por los componentes de hardware y software de un dispositivo móvil.

A partir de la investigación realizada surgirá una propuesta para la utilización de dispositivos y redes móviles en zonas rurales aisladas, en la misma se describirán técnicas y recomendaciones para reducir el consumo de energía de un dispositivo móvil y estrategias para recargar las baterías utilizando energía solar fotovoltaica.

Desde una perspectiva social, las principales contribuciones que se esperan de la propuesta son:

- Disminución de la brecha digital existente entre las poblaciones rurales aisladas y las poblaciones urbanas.
- Reducción de la migración de los pobladores de zonas rurales aisladas a zonas urbanas.
- Aprovechamiento de la energía solar fotovoltaica para la recarga de baterías de dispositivos portátiles en zonas aisladas. Los pobladores o visitantes de estas zonas no tendrán necesidad de desplazarse a centros urbanos para recargar sus equipos.
- Mejoras en el aprovechamiento de la tecnología disponible en zonas aisladas, teniendo en cuenta que

algunos de los pobladores de estas zonas son propietarios de equipos celulares que utilizan como reproductores de música o cámaras fotográficas y no como dispositivos de comunicación.

- Incremento de las posibilidades de comunicación de los pobladores de la zona rural, permitiendo el acceso a aplicaciones de Internet tales como correo electrónico, mensajería y redes sociales.
- Mejoras en el proceso de enseñanza aprendizaje en comunidades escolares rurales aisladas, mediante la implementación de estrategias educativas de m-learning.
- Fomento del turismo rural, brindando servicios de comunicación a zonas aisladas que formen parte de los circuitos turísticos del NOA.

Los resultados obtenidos por el grupo de investigación fueron publicados en [9] [10] [11] [12].

5 Formación de recursos humanos

La propuesta involucra la integración de los conocimientos en esta área por parte de investigadores de la Universidad Nacional de Salta (UNSa) y de la Universidad Nacional de Santiago del Estero (UNSE). El año 2015 se presentó una tesis de postgrado en el Magister en Redes de la UNLP [12]. Uno de los investigadores se encuentra realizando el doctorado en Ciencias Informáticas de la UNLP, el trabajo de investigación de la propuesta de tesis [13] se encuentra directamente relacionado con este proyecto.

El proyecto cuenta con la participación de estudiantes avanzados de carreras de grado en Informática, pertenecientes a la UNSa y a la UNSE, los cuales reciben formación en las áreas de computación móvil y energía solar; además, experiencia en el desarrollo de investigaciones. Dos alumnos están finalizando la tesis de grado de la Licenciatura en Sistemas de Información de la Facultad de Facultad de Ciencias Exactas y Tecnologías de la UNSE, y otro se encuentra desarrollando el seminario de sistemas de la Licenciatura en Análisis de Sistemas de la Facultad de Ciencias Exactas de la UNSa.

Referencias

1. Ottavianelli, Emilce; Ibarra Marcelo; Cadena Carlos. (2013). *Uso de indicadores sociales en estudios de factibilidad de instalación de sistemas solares para generación de electricidad en localidades rurales*. Paper presented at the XX Simposio peruano de Energía solar, Tacna – Perú.
2. Secretaria de Energia - Ministerio de Planificación. (2015). PERMER, Proyecto de Energías Renovables en Mercados Rurales. from <https://www.se.gob.ar/permer/>
3. Thiagarajany, Narendran; Aggarwal, Gaurav; Nicoara, Angela (2012). *Who Killed My Battery*:

- Analyzing Mobile Browser Energy Consumption*. Paper presented at the WWW 2012 – Session: Mobile Web Performance, Lyon, France.
4. ZHANG, Lide; TIWANA, Birjodh; QIAN, Zhiyun and WANG, Zhaoguang. (2010). *Accurate online power estimation and automatic battery behavior based power model generation for smartphones*. Paper presented at the 2010 IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS), Scottsdale, AZ. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5751489>
 5. Carroll, Aaron; Heiser, Gernot (2013). *The systems hacker's guide to the galaxy energy usage in a modern smartphone*. Paper presented at the 4th Asia-Pacific Workshop on Systems (APSys '13).
 6. Metri, Grace; Shi, Weisong; Brockmeyer Monica (2015). *EnergyEfficiency Comparison of Mobile Platforms and Applications: A Quantitative Approach*. Paper presented at the HotMobile Santa Fe, NM, USA.
 7. ARDUINO UNO Board. (2015). from <https://www.arduino.cc/en/Main/arduinoBoardUno>
 8. Nupur K. Sonawane. (2014). Bluetooth Based Device Automation System Using Cellphone. *International Journal of Computer Applications & Information Technology, Vol. 7 Issue 1*.
 9. Rocabado, Sergio; Cadena, Carlos. (2015). *Despliegue de MANETs en zonas rurales aisladas*. Paper presented at the WICC 2015.
 10. Rocabado, Sergio; Cadena, Carlos. (2015). *Uso de dispositivos móviles inteligentes en zonas rurales aisladas abastecidas con energía solar fotovoltaica*. Paper presented at the CACIC 2015.
 11. Rocabado, Sergio; Cadena, Carlos. (2015). *Cargadores solares portátiles para el uso de dispositivos móviles en zonas rurales aisladas del NOA*. Paper presented at the ASADES 2015.
 12. Rocabado, Sergio; Cadena, Carlos. (2016). *Mini sistemas fotovoltaicos para el uso de dispositivos móviles en zonas rurales: Optimización de la potencia entregada y consumida*. Paper presented at the Congreso Brasileiro de Energía Solar - CBENS 2016.
 13. Rocabado, Sergio. (2015 - 2017). *Optimización del consumo energético en dispositivos móviles para su uso en zonas rurales aisladas abastecidas con energía solar fotovoltaica*. Doctorado en Ciencias Informáticas. UNLP.
 14. Tarkoma, Sasu; Siekkinen, Matti; Lagerspetz, Eemil; Xiao, Yu (2014). *Smartphone Energy Consumption: Modeling and Optimization* (First ed., pp. 234). University Printing House, Cambridge CB2 8BS, United Kingdom: Cambridge University Press. E-book ISBN: 9781139990417.
 15. Attia, Y. I. Al-Mashhadany; H. A. (2014). Novel Design and Implementation of Portable Charger through Low-Power PV Energy System. *Advanced Materials Research, 925*, pp. 495-499 (ISSN 1662-8985).
 16. Attia, Yousif I. Al-Mashhadany; Hussain A. (2014, November 12). *High performance for real portable charger through lowpower PV system*. Paper presented at the International Journal of Sustainable and Green Energy.
 17. Qualcomm Technologies. (2014). Trepp Profiler. from <https://developer.qualcomm.com/mobile-development/increase-app-performance/trepp-profiler>
 18. Zhang, Lide. (2013). *Power, Performance Modeling and Optimization for Mobile System and Applications*. (Thesis - Doctor of Computer Science and Engineering), University of Michigan.
 19. Yevgen, Barsukov; Jinrong, Qian. (2013). *Battery Power Management for Portable Devices* (First ed., pp. 268). Norwood, MA 02062 USA: Artech House. E-book ISBN: 978-1-60807-491-4.
 20. Schuss, C; Rahkonen, T. (2013, 11-15 Nov.). *Solar Energy Harvesting Strategies for Portable Devices such as Mobile Phones*. Paper presented at the 2013 14th Conference of Open Innovations Association (FRUCT) - IEEE, Espoo.
 21. Michael, Boxwell. (2015). *Solar Electricity Handbook : A simple, practical guide to solar energy - designing and installing solar PV systems* (2015 ed., pp. 204). Greenstream Publishing. E-book ISBN: 978-1907670459.
 22. Távora, Filipe; Maia, Andrea Sarmiento. (2012). *Solar Batiery charger for portable devices application*. SiliconReef Consultoria, Pesquisa e Projetos em Tecnologia da Informação. Retrieved from www.siliconreef.com.br
 23. Carroll, Aaron; Heiser, Gernot (2010). *An Analysis of Power Consumption in a Smartphone*. Paper presented at the USENIX Annual Technical Conference, Boston, MA, USA.
 24. Salgado Fernandez, Jose (2010). *Compendio de energia solar. Fotovoltaica Termica Termoelectrica*. A. Madrid Vicente - Mundi Prensa. E-book ISBN: 9788484764007.

Redes de Sensores Inalámbricos en Entornos Industriales para monitorear Condiciones de Higiene y Seguridad

Eduardo Rodríguez, Claudia Deco, Luciana Burzacca, Mauro Pettinari,
Santiago Costa, Cristina Bender

Departamento de Investigación Institucional,
Facultad de Química e Ingeniería del Rosario,
Universidad Católica Argentina Campus Rosario
Avda. Pellegrini 3314, S2002QEO, Rosario, Santa Fe
{ejrodriguez, cdeco, lburzacca, mauro_pettinari, santiagocosta, cbender}@uca.edu.ar

Resumen

Las redes de sensores inalámbricos consisten en una serie de nodos equipados con capacidades de procesamiento, comunicación y sensado. Utilizan protocolos especiales de radio para transmitir datos en un modo multisalto de operación. La medición de parámetros físicos hace que los sensores sean la tecnología más adecuada para el monitoreo y reporte de mediciones cuantificables. En este proyecto se propone utilizar una red de sensores para el monitoreo de las condiciones ambientales de Higiene y Seguridad en entornos industriales. En este proyecto se monitorean temperatura, humedad, ruido y luminosidad, con el fin de dar soporte a la inspección anual de un auditor externo, por lo que no se considera esta recolección como crítica dado que no controlan ningún dispositivo. En primera instancia se aborda el problema utilizando una red de sensores con una plataforma Arduino anexada a módulos Xbee. En una segunda instancia se aborda este problema con el enfoque de Internet de las Cosas (IoT), utilizando módulos Arduino con WiFi embebido, denominados Wido, que permiten la comunicación de datos directamente al servidor de almacenaje.

Palabras Claves: Redes de Sensores, Monitoreo Condiciones de Higiene y Seguridad, Entornos Industriales, Internet de las Cosas.

Contexto

Esta línea de I+D se está llevando a cabo a través de proyectos del Departamento de Investigación Institucional de la Facultad de Química e Ingeniería del Rosario de la Universidad Católica Argentina.

Los proyectos involucrados son: PID UCA “Optimización de Redes Inalámbricas de uso Comunitario” (2011 - 2014); y PID UCA “Monitoreo y Análisis de Condiciones de Higiene y Seguridad en Entornos Industriales usando Redes de Sensores Inalámbricos” (2014 - 2017).

Introducción

El objetivo del monitoreo de condiciones ambientales es una recolección de información eficiente, que se utiliza tanto para la prevención (en tiempo real o pospuesto) como para el análisis. La migración de las redes de sensores cableadas a las redes inalámbricas trae numerosas ventajas al facilitar el proceso de despliegue y la recolección de la información. De hecho,

debido a su pequeño tamaño, los sensores pueden ser fácil y rápidamente desplegados en grandes escalas a bajo costo. Sus funciones inalámbricas los hacen independientes de cualquier infraestructura costosa y fija y también contribuyen a su éxito.

Una red de sensores inalámbrica (Wireless Sensor Network o WSN) se caracteriza por su facilidad de despliegue y por ser auto-configurables, pudiendo un nodo convertirse en todo momento en emisor, receptor, ofrecer servicios de encaminamiento entre nodos sin visión directa, así como registrar datos referentes a los sensores locales de cada nodo. Otra de sus características es su gestión eficiente de la energía, que les permite obtener una alta tasa de autonomía que las hacen plenamente operativas.

La evolución tecnológica trajo aparejado equipamiento de pequeño tamaño con capacidad de procesamiento, conectividad inalámbrica y de bajo costo. Estos dispositivos se denominan motes con una unidad de procesamiento de cómputo mínimo, memoria, una unidad de comunicación inalámbrica y uno o varios dispositivos de sensado que capturan parámetros como temperatura, humedad, etc. Una red de sensores inalámbrica está formada por un conjunto de motes comunicados entre sí. Su distribución puede ser aleatoria o planificada. Los nodos pueden trabajar de modo cooperativo [1, 2].

Las aplicaciones de redes de sensores inalámbricos en entornos industriales ha crecido mucho en los últimos años y prueba de ello son los distintos trabajos de investigación y artículos existentes [3, 4, 5, 6, 7].

Si bien desde la concepción del proyecto se pensó en una red de sensores inalámbricos para abordar el problema, posteriores avances nos llevan a la necesidad de contrastar la propuesta con

un enfoque de Internet de las Cosas (IoT), a los fines de evaluar los costos de una y otra solución.

Se denomina Internet de las Cosas al concepto que define la red global de información y comunicación en donde todos los objetos que nos rodean, independientemente de su naturaleza, tamaño y geometría, se encuentran identificados y conectados permanentemente a Internet. Esto permite la captura, el almacenamiento y la gestión de toda la información emitida por dichos objetos con la finalidad de automatizar actividades y procesos diarios en la vida cotidiana así como analizar todos los datos generados aportando información útil que ayude a la correcta toma de decisión frente a determinadas situaciones [8]. Las tecnologías de software y hardware en las que se apoya están totalmente desarrolladas y listas para implantarse. Por ejemplo, tecnologías como Big Data, Business Intelligence, Analytics, Cloud Computing, dispositivos Wearables, etiquetas RFID, fibra óptica, comunicaciones Wireless, Smart cities, etc. Una ventaja es la capacidad de poder ser aplicada por ejemplo a sensores ubicados en múltiples puntos recogiendo información sobre parámetros ambientales.

El proyecto comprende la caracterización del problema, la elección del hardware, el diseño de la red y la realización de pruebas para demostrar su funcionamiento. La recolección de estos datos no se considera crítica ya que no controlan ningún dispositivo y se propone esta recolección de datos para dar soporte a la inspección anual de un auditor externo. Esto nos lleva a que para solucionar el problema se pueden utilizar protocolos y dispositivos que no están necesariamente diseñados para entornos industriales [9]. En este proyecto se monitorean Temperatura, Humedad,

Ruido y Luminosidad. Los valores límites de estos parámetros y la forma de realizar estos controles son los establecidos por la Ley N° 19587 de Higiene y Seguridad y reglamentaciones correspondientes [10]. Esta ley fue consultada a los efectos de tener en cuenta los criterios y valores para Temperatura y Humedad (índice TGBH), Ruido (valores de dBA) y Luminosidad (unidad de medida LUX) establecidos.

Si bien existen trabajos de investigación en el área para entornos industriales no conocemos específicamente trabajos aplicados al área de higiene y seguridad general en estos entornos. Dada la importancia de este tema en la industria creemos que nuestro proyecto es un aporte significativo a la utilización de redes de sensores en dichos entornos. La realización de este proyecto permite traducir el conocimiento generado en investigación en un importante desarrollo tecnológico para un sector de alta relevancia como es el de la industria.

Líneas de Investigación, Desarrollo e Innovación

En este proyecto, se utilizan dos abordajes para resolver el problema: uno usando redes de sensores inalámbricos y otro según el enfoque utilizado en Internet de las cosas. En ambos casos se emplean plataformas Arduino o compatibles que manejan la recolección de los datos provenientes de sensores que se anexan a las mismas. Para la transmisión de los datos existen distintos protocolos. En particular, en este proyecto se utiliza el protocolo MQTT (MQ Telemetry Transport, <http://mqtt.org/>) con el Servidor Mosquitto, dado que este protocolo se encuentra disponible para plataformas Arduino y puede ser utilizado en ambos abordajes. MQTT es un

protocolo de transporte de mensajes en la forma Cliente/Servidor. Es liviano, abierto, sencillo y está diseñado para que sea fácil de implementar. Estas características lo hacen ideal para su uso en muchas situaciones, incluyendo entornos limitados como para la comunicación de máquina a máquina (M2M) y el Internet de las Cosas (IoT).

Para el enfoque de Redes de Sensores se utilizan módulos inalámbricos Xbee los cuales están conectados a los Arduino recibiendo los datos de ellos y transmitiéndolos a los otros nodos. La red que forman los Xbee es una red mallada que una vez configurada se encarga de transmitir los datos que recibe cada uno de los nodos a un nodo central. En dicha red existen tres tipos de nodos: Coordinador, Router y End Device. El Coordinador es responsable de establecer el canal de comunicaciones y del PAN ID (identificador de red) para toda la red. Una vez establecidos estos parámetros, el Coordinador puede formar una red, y participa en el enrutado de paquetes y es origen y/o destinatario de información. El Router crea y mantiene información sobre la red para determinar la mejor ruta para enrutar un paquete de información. Y los End Device son los dispositivos finales no tienen capacidad de enrutar paquetes. La arquitectura de esta red es sencilla y puede observarse en la Figura 1, donde C es el nodo Coordinador, R son los nodos Routers y E son los dispositivos finales.

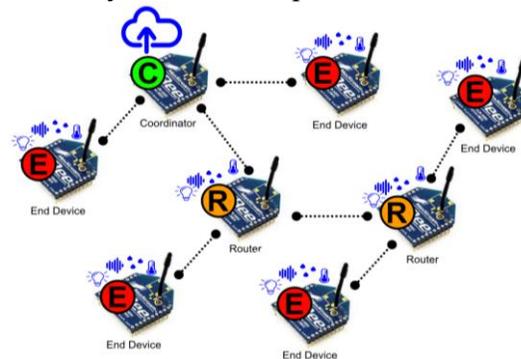


Figura 1 *Arquitectura Red de sensores.*

Desde el punto de vista de Internet de las Cosas (IoT) se utilizan módulos Arduino compatibles denominados Wido que tienen un módulo WiFi embebido. Estos dispositivos toman los datos de los sensores de la misma manera que en el abordaje anterior y los transmiten directamente al host donde se encuentra el servidor que recoge los datos y los almacena en una base de datos. Para lograr esta transmisión el Wido se conecta a un Punto de Acceso inalámbrico o a una red mallada que está conectada a Internet. La arquitectura de esta red también es sencilla y puede observarse en la Figura 2.

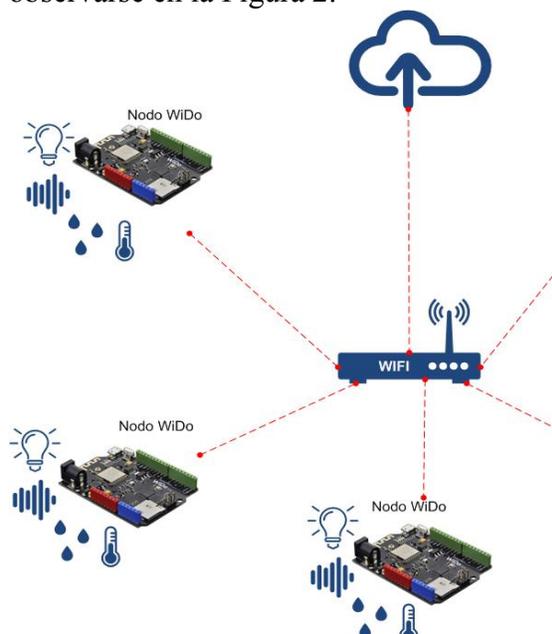


Figura 2 Arquitectura Red IOT.

Resultados y Objetivos

En este proyecto se trata la implementación de un sistema de sensado y monitoreo permanente de condiciones ambientales en entornos industriales, lo cual redundará en beneficio de las condiciones laborales de los operarios. De todas maneras dado que la ley no exige este tipo de monitoreo su implementación depende de la voluntad de la empresa.

Se presentan dos tipos de abordaje

para el problema. En los primeros análisis para hacer una estimación de la cantidad de datos a manejar, se calculó una frecuencia de muestreo de 1 minuto, por lo que se tienen 4 datos por minuto por cada nodo, lo que implica 5760 datos diarios por nodo. En este proyecto se trabajó con 10 nodos pero una aplicación real debe ser capaz de manejar cientos de ellos. Un cálculo estimativo del costo para el abordaje con redes de sensores, es de 210 dólares por nodo; y para el abordaje de IoT, este costo estimativo es de 130 dólares por nodo. Con lo cual, implementar una red IoT para 10 nodos, resulta un 65% del costo de implementar una red de sensores para la misma cantidad de nodos [11].

Actualmente, siguiendo la línea de IoT, se está realizando la evaluación del reemplazo de los Wido por dispositivos similares denominados NodeMCU. Estos dispositivos de reciente aparición tienen un costo mucho menor y eso haría la propuesta mucho más atractiva para la industria. En este momento un Wido cuesta U\$S 25 y un NodeMCU U\$S 5, si las pruebas de desempeño son satisfactorias se tendría una reducción de costo significativa para la misma cantidad de nodos.

La evaluación de los resultados obtenidos se está realizando mediante un análisis cuantitativo y cualitativo de las propuestas en comparación con la utilización de sensores cableados tradicionales o mediciones esporádicas con elementos manuales. En este sentido, se valida si los datos obtenidos del monitoreo y el comportamiento de la red experimental son relevantes y dan cuenta de si este abordaje puede reemplazar al tradicional y cuáles son sus principales ventajas.

La conformación del grupo de investigación por docentes de la carrera de Ingeniería Industrial y de la carrera de Licenciatura en Sistemas permite un

enfoque multidisciplinario y adecuado de esta problemática. La viabilidad del proyecto está sustentada en los avances obtenidos mediante trabajos previos de los integrantes en el área y su interacción con otros grupos de investigadores tanto latinoamericanos como europeos.

Se espera que los resultados obtenidos en el proyecto ayuden y/o mejoren la implementación del monitoreo de condiciones de higiene y seguridad en la industria mejorando significativamente la condiciones de trabajo del personal al estar éstas controladas en forma más efectiva.

Formación de Recursos Humanos

El equipo de trabajo está integrado por los Magisters Eduardo Rodríguez y Cristina Bender, la Doctora Claudia Deco, la especialista Luciana Burzacca, y los Licenciados Santiago Costa y Mauro Pettinari, investigadores de la Universidad Católica Argentina.

En particular, el Lic. Santiago Costa está cursando la Maestría en Redes de Datos en la Universidad de La Plata

Además, los conocimientos adquiridos se volcarán en la comunidad académica a través de las actividades docentes que los investigadores de este proyecto realizan en las carreras Ingeniería Industrial, Ingeniería Ambiental y Licenciatura en Sistemas. Asimismo, es intención incorporar alumnos de estas carreras y proponer subproblemas identificados durante el trabajo de investigación de este proyecto para la realización de tesinas y proyectos finales de grado.

Referencias

[1] D. Culler and W. Hong. Eds. Special issue on Wireless Sensor Networks. *Communications of the ACM*, 47(6):pag 30–34, June 2004

- [2] G. Mercado, A. Diedrichs y M. Aguirre “The Wireless Embedded Internet”, *Annals of CASE 2011*, ISBN 978-987-9374-69-6, Buenos Aires Marzo 2011.
- [3] D. Culler; D. Estrin, M. Srivastava: “Overview of Sensor Networks” *IEEE Computer Mag.* 2004, 37, 41–49
- [4] S Yoo, PK Chong, D Kim, Y Doh. Guaranteeing real-time services for industrial wireless sensor networks with IEEE 802.15. 4. *Industrial Electronics*, 2010 – ieeexplore.ieee.org
- [5] V. Güngör, B. Lu, G. Hancke. Opportunities and challenges of wireless sensor networks in smart grid. *Industrial Electronics. IEEE Transactions on*, 57 (10): 3557-3564. Oct 2010
- [6] J Åkerberg, M Gidlund, T Lennvall, J Neander. Efficient integration of secure and safety critical industrial wireless sensor networks. *Journal on Wireless*, Springer, 2011
- [7] V. Güngör; G. Hancke *Industrial Wireless Sensor Networks: Applications, Protocols, and Standards*. CRC Press. April 04, 2013 - 978-1-4665-0051-8
- [8] Conner, Margery. Sensors empower the "Internet of Things". *EDN Network (Issue 10)*. pp. 32–38. Disponible en www.edn.com/design/sensors/4363366/Sensors-empower-the-quot-Internet-of-Things-quot-. 2010. ISSN 0012-7515
- [9] V. Güngör, V. Çağrı; G. Hancke. *Industrial Wireless Sensor Networks: Applications, Protocols, and Standards*. CRC Press. April 4, 2013.
- [10] R. A. Parada (comp.) y J. D. Errecaborde (comp.). *Higiene y seguridad en el trabajo. Versión 2.1. ERREPAR S.A.*. Páginas 7-92. Buenos Aires. ISDN 978-987-01-1750-6. 2015.
- [11] E. Rodríguez, C. Deco, L. Burzacca, M. Pettinari, S. Costa, C. Bender. Análisis y diseño de una red sensores con tecnología inalámbrica para el monitoreo de condiciones de Higiene y Seguridad del ambiente en entornos industriales. *Congreso Argentino Ingeniería Industrial. Facultad Regional Córdoba, UTN Córdoba, Argentina. Noviembre 2015.*

Arquitecturas adaptadas para integrar computación móvil y computación en la nube

Karina Cenci

Leonardo de- Matteis

Jorge R. Ardenghi

Laboratorio de Investigación en Sistemas Distribuidos

Universidad Nacional del Sur

Bahía Blanca - Buenos Aires - Argentina

e-mail: {kmc,ldm,jra}@cs.uns.edu.ar

Resumen

Hoy en día, el concepto de computación en la nube y sus servicios asociados comienza a ser una realidad y su uso es cada vez más frecuente en diversos ámbitos, tanto privados como públicos. Pero la adopción de la nube, como infraestructura de almacenamiento o aplicaciones, es un poco lenta a nivel de las organizaciones, principalmente debido a aspectos relacionados con tres factores: las políticas de confidencialidad de la información, la problemática de una disponibilidad de acceso confiable/permanente y la operabilidad sin interrupciones de los servicios implementados sobre la nube. En nuestro laboratorio planteamos una serie de objetivos para el estudio y realización de trabajos relacionados con: movilidad de dispositivos actuales, estrategias adecuadas para mantener consistencia de datos ante conexiones/desconexiones frecuentes, migración y distribución tanto de datos como de servicios, confección y estudio de nuevas propuestas para una nube híbrida que permita abarcar en forma adecuada la problemática asociada a la privacidad.

Palabras claves: Sistemas Distribuidos, Computación en la nube, Movilidad.

Contexto

Esta línea de investigación y desarrollo se lleva adelante en el ámbito del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur (UNS), en particular, como parte de las tareas que se realizan en el Laboratorio de Investigación en Sistemas Distribuidos (LISiDi). El proyecto se financia parcialmente con fondos asignados por la UNS.

Introducción

La evolución de la computación comienza con las primeras computadoras *mainframe* con modelos centralizados, luego se evoluciona y comienzan a ganar preferencia las personales, migrando a un nuevo esquema basado en el modelo cliente-servidor. Más recientemente se constata el surgimiento de nuevas tecnologías para la interconexión global y se ingresa en la era de Internet. Actualmente, las organizaciones y los usuarios pueden comunicarse a través de una red de computadoras de área amplia, incorporando progresivamente dispositivos móviles dispersos en el ambiente.

La computación en la nube (*Cloud Computing*) surge naturalmente como un modelo para permitir en forma conveniente y bajo demanda el acceso a la red sobre un conjunto de recursos informáticos compartidos personalizables que pueden ser rápidamente aprovisionados y liberados con un mínimo esfuerzo de gestión o de interacción del proveedor de servicios. La computación en la nube es cada vez más popular como un medio para ejecutar grandes programas en paralelo sobre una colección de equipos organizados en un grid. La construcción de un sistema que permita a miles de computadoras distribuidas por Internet ser compartidas por los usuarios requiere una mirada integral para un adecuado desarrollo, teniendo en cuenta aspectos de escalabilidad, seguridad, disponibilidad, tolerancia a fallas, sistemas operativos, apoyo a la programación paralela, servicios de búsqueda, administración de sistemas, reserva de recursos, heterogeneidad y muchos más.

A su vez, los dispositivos móviles se están convirtiendo en omnipresentes, con interfaz universal a los servicios y aplicaciones en línea. Más recién-

temente el paradigma denominado Internet de las cosas (*Internet of Things*), está experimentando un crecimiento exponencial. Bajo este esquema, los dispositivos embebidos de pequeñas dimensiones y con recursos limitados forman nuevas redes y se conectan a las existentes, compartiendo información con otros dispositivos o servicios. Sin embargo, la capacidad de ejecutar aplicaciones con altos requerimientos se ve reducida dada la limitada capacidad de procesamiento, almacenamiento y/o autonomía energética que poseen estos dispositivos.

Por otra parte, se introduce una necesidad cada vez mayor con respecto al acceso a los diferentes tipos de documentos o archivos con los cuales trabaja el usuario diariamente desde cualquier dispositivo en cualquier lugar donde se encuentre. Pero el usuario todavía no cuenta con un mecanismo que le permita tener todos sus archivos en la nube, a veces por cuestiones de espacio, o bien por desconocimiento de tecnologías y otras por falta de confianza en mecanismos que aseguren la privacidad de los datos almacenados en servidores de terceros que brindan dichos servicios globales (Onedrive, Dropbox, Gdrive, etc.)

Recientemente, han surgido alternativas basadas en protocolos P2P para la distribución de los archivos personales sobre todo el conjunto de dispositivos con los cuales cuenta el usuario. Tales alternativas permiten alcanzar características deseables como privacidad, confiabilidad y autenticidad y no dependen de servicios provistos por terceros, sino que el mismo usuario con sus dispositivos conforma una nube privada propia (Syncthing, BitTorrent Sync).

También disponemos, en la actualidad, de dispositivos móviles que incluyen la característica de operación con perfiles duales (corporativo y privado del usuario). Entonces es admisible incorporar esta dualidad sobre la nube en sí misma, surgiendo así el concepto de nube híbrida, la que será accesible por los usuarios y controlada parcialmente por las políticas establecidas en el ámbito institucional o empresarial de cada organización, respetando de este modo la privacidad del espacio asignado al usuario como individuo y sin que este, al utilizarlo, afecte a la seguridad o contradiga las políticas preestablecidas sobre el espacio de uso corporativo.

Líneas de Investigación y Desarrollo

Los estudios sobre los sistemas distribuidos y paralelos, los paradigmas de computación en la nube, las aplicaciones basadas en protocolos Peer-to-Peer y la computación móvil han adquirido gran importancia en aplicaciones sociales como son los sistemas de información y comunicación (*home-banking*, servicio de compras en línea, *dashboards* ejecutivos para análisis de datos y seguimiento de procesos).

Los resultados que se obtengan se pueden extender a aplicaciones basadas en servicios (*Service Oriented Architecture - SOA*), al desarrollo de aplicaciones distribuidas de gobierno electrónico y al empleo de documentos y datos en pequeña y mediana escala.

El contexto de estos desarrollos se puede aplicar en organizaciones y empresas de pequeña y mediana escala para mejorar los procesos del negocio y las políticas de gobierno. El acceso a documentos y a bases de datos desde diferentes locaciones fuera de los dominios de una organización es altamente requerido y está en continuo crecimiento. Favorece el diseño de las aplicaciones SOA y, además, la facilidad de brindar servicios que estén distribuidos y que promuevan un mejor aprovechamiento de los recursos de los equipos y reduzcan el tráfico de red.

Resultados y Objetivos

A partir de las necesidades que surgen en las organizaciones por el incremento de dispositivos y conectividad, factores asociados a los que miembros de los equipos de trabajo realizan sus tareas en diferentes locaciones, el acceso a documentos y datos que se encuentran dentro de la organización resulta indispensable. En [6] y [7], se propone una arquitectura basada en capas como una alternativa para acceder a documentos compartidos. La evolución del trabajo da origen al desarrollo de este proyecto que se encuentra en su etapa inicial.

Los objetivos generales que persigue en este proyecto son:

- Aprovechamiento de la movilidad de los dispositivos.
- Migración de datos y servicios para reducir la latencia en el tiempo de acceso. Una pla-

taforma distribuida en el borde de la nube que desacopla los servicios desde lugares de alojamiento fijos para reducir la latencia de acceso a los servicios de back-end - mediante la migración de los servicios en tiempo de ejecución para adaptarse a los cambios en las condiciones de la red debido a la itinerancia del usuario.

- Habilidad para la conexión y desconexión de la red, garantizando el estado de consistencia.
- Distribución de servicios y datos entre diferentes nodos. Se considera especialmente la sincronización de archivos de las cuentas de los usuarios entre los dispositivos, por un lado, para identificar dónde se almacenan los documentos y alcanzar mayor grado de privacidad y, por otra parte, se tiene en cuenta cómo los documentos pueden ser accedidos y modificados a pesar de no contar con servicios de Internet.
- Investigación y análisis de cómo aprovechar conceptos y características de sistemas tipo EMM (*Enterprise Mobility Management*) utilizados en ámbitos corporativos e institucionales (BlackBerry Enterprise Server, MobileIron, etc.) para el agregado de nuevas funcionalidades de computación en la nube.

Formación de recursos humanos

En relación con la formación de los recursos humanos, el trabajo a desarrollar permitirá una capacitación práctica adecuada de los profesionales involucrados, favoreciendo que alcancen logros concretos a partir de las investigaciones realizadas. Además, esta línea de investigación permitirá la dirección de tesis de licenciatura y trabajos finales de ingeniería afines. Asimismo, y por último, se podrá generar *courseware* para materias optativas a dictarse en el Departamento de Ciencias e Ingeniería de la Computación de la UNS.

Referencias

- [1] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, et al. A view of cloud computing. *Communications of the ACM*, 53(4):50–58, 2010.
- [2] Greg Boss, Padma Malladi, Dennis Quan, Linda Legregni, and Harold Hall. Cloud computing. *IBM white paper*, 321:224–231, 2007.
- [3] Gerard Briscoe and Alexandros Marinos. Digital ecosystems in the clouds: towards community cloud computing. In *Digital Ecosystems and Technologies, 2009. DEST'09. 3rd IEEE International Conference on*, pages 103–108. IEEE, 2009.
- [4] Rajkumar Buyya, James Broberg, and Andrzej M Goscinski. *Cloud computing: principles and paradigms*, volume 87. John Wiley & Sons, 2010.
- [5] Rajkumar Buyya, Chee Shin Yeo, and Srikumar Venugopal. Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities. In *High Performance Computing and Communications, 2008. HPCC'08. 10th IEEE International Conference on*, pages 5–13. Ieee, 2008.
- [6] Karina M Cenci, Leonardo de Matteis, and Jorge Raúl Ardenghi. Arquitectura en capas para acceso remoto sad. In *XVIII Congreso Argentino de Ciencias de la Computación*, 2013.
- [7] Karina M Cenci, Leonardo de Matteis, and Jorge Raúl Ardenghi. Tiered architecture for remote access to data sources. *Journal of Computer Science & Technology*, 14, 2014.
- [8] S Chetan, Gautam Kumar, K Dinesh, K Mathew, and MA Abhimanyu. Cloud computing for mobile world. *available at chetan. ueuo. com*, 2010.
- [9] Byung-Gon Chun, Sunghwan Ihm, Petros Maniatis, Mayur Naik, and Ashwin Patti. Clonecloud: elastic execution between mobile device and cloud. In *Proceedings of the sixth conference on Computer systems*, pages 301–314. ACM, 2011.
- [10] Gang Ding and Bharat Bhargava. Peer-to-peer file-sharing over mobile ad hoc networks. In *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, pages 104–108. IEEE, 2004.
- [11] Amy Elser. *Guide to reliable distributed systems: building high-assurance applications and cloud-hosted services*. Springer Science & Business Media, 2012.
- [12] Niroshinie Fernando, Seng W Loke, and Wenny Rahayu. Mobile cloud computing: A survey. *Future Generation Computer Systems*, 29(1):84–106, 2013.
- [13] Guido Gehlen and Linh Pham. Mobile web services for peer-to-peer applications. In *Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE*, pages 427–433. IEEE, 2005.

- [14] Ioana Giurgiu, Oriana Riva, Dejan Juric, Ivan Krivulev, and Gustavo Alonso. Calling the cloud: enabling mobile phones as interfaces to cloud applications. In *Middleware 2009*, pages 83–102. Springer, 2009.
- [15] Péter Kacsuk, Róbert Lovas, and Zsolt Németh. *Distributed and Parallel Systems: In Focus: Desktop Grid Computing*. Springer Science & Business Media, 2008.
- [16] Matthias Kovatsch, Martin Lanter, and Zach Shelby. Californium: Scalable cloud services for the internet of things with coap. In *Internet of Things (IOT), 2014 International Conference on the*, pages 1–6. IEEE, 2014.
- [17] Matthias Kovatsch, Simon Mayer, and Benedikt Ostermaier. Moving application logic from the firmware to the cloud: Towards the thin server architecture for the internet of things. In *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on*, pages 751–756. IEEE, 2012.
- [18] Ronald L Krutz and Russell Dean Vines. *Cloud security: A comprehensive guide to secure cloud computing*. Wiley Publishing, 2010.
- [19] Yan Lu, Shipeng Li, and Huifeng Shen. Virtualized screen: A third element for cloud mobile convergence. *MultiMedia, IEEE*, 18(2):4–11, 2011.
- [20] Zaigham Mahmood and Richard Hill. *Cloud Computing for enterprise architectures*. Springer Science & Business Media, 2011.
- [21] Lijun Mei, Wing Kwong Chan, and TH Tse. A tale of clouds: paradigm comparisons and some thoughts on research issues. In *Asia-Pacific Services Computing Conference, 2008. APSCC'08. IEEE*, pages 464–469. Ieee, 2008.
- [22] Johann Van Der Merwe, Dawoud Dawoud, and Stephen McDonald. A survey on peer-to-peer key management for mobile ad hoc networks. *ACM computing surveys (CSUR)*, 39(1):1, 2007.
- [23] Siani Pearson. Privacy, security and trust in cloud computing. In *Privacy and Security for Cloud Computing*, pages 3–42. Springer, 2013.
- [24] Jan S Rellermeyer, Gustavo Alonso, and Timothy Roscoe. R-osgi: distributed applications through software modularization. In *Proceedings of the ACM/IFIP/USENIX 2007 International Conference on Middleware*, pages 1–20. Springer-Verlag New York, Inc., 2007.
- [25] Gautam Shroff. *Enterprise cloud computing: technology, architecture, applications*. Cambridge University Press, 2010.
- [26] Ziyuan Wang. Security and privacy issues within the cloud computing. In *Computational and Information Sciences (ICCIS), 2011 International Conference on*, pages 175–178. IEEE, 2011.
- [27] Lamia Youseff, Maria Butrico, and Dilma Da Silva. Toward a unified ontology of cloud computing. In *Grid Computing Environments Workshop, 2008. GCE'08*, pages 1–10. IEEE, 2008.
- [28] Xinwen Zhang, Anugeetha Kunjithapatham, Sangoh Jeong, and Simon Gibbs. Towards an elastic application model for augmenting the computing capabilities of mobile devices with cloud computing. *Mobile Networks and Applications*, 16(3):270–284, 2011.

“Análisis de Eficiencia en Arquitecturas Multiprocesador para Aplicaciones de Transmisión y Procesamiento de Datos”

Jorge R. Osio^{1,2}, Juan E. Salvatore¹, Eduardo Kunysz¹, Valentin Guarepi¹,
Daniel Martin Morales^{1,3}

¹Instituto de Ingeniería y Agronomía - Universidad Nacional Arturo Jauretche

²UIDET CeTAD – Fac. de Ingeniería - Universidad Nacional de La Plata

³Laboratorio LINES - Universidad Tecnológica Nacional FRLP

josio@unaj.edu.ar, jsalvatore@unal.edu.ar, ekunys@unaj.edu.ar, vguarepi@unaj.edu.ar,
martin.morales@unaj.edu.ar

Resumen

Los principales lineamientos de investigación de este proyecto se orientan a las Arquitecturas que posibilitan el procesamiento paralelo, principalmente en sistemas multi-core.

Por otro lado, la optimización de estos sistemas se orienta a determinadas aplicaciones de telemedicina y de sistemas ópticos basados en MIOC , (Multifunction Integrated Optical Chip), que requieren procesamiento y transmisión de datos de forma remota.

Paralelamente a estos lineamientos el proyecto contempla el estudio de las nuevas tendencias en cómputo paralelo con el objetivo de mejorar rendimiento de cálculo. Como una alternativa interesante para optimizar la relación costo-prestaciones, se exploran tecnologías de computadoras de alta performance reconfigurable (HPCR).

En relación a la comunicación Ethernet, se investiga principalmente el protocolo UDP para la transmisión de imágenes y datos de audio y video, donde no es

crítica la pérdida de algún que otro dato, pero es importante el envío de gran caudal de datos en tiempo real. Adicionalmente, para los pequeños paquetes de datos críticos se utiliza TCP.

Palabras clave: *Arquitecturas Paralelas, Telemedicina, Software Embebido, Sistemas Multicore, Platadormas FPGAs.*

Contexto

Las líneas de Investigación aquí desarrolladas son parte del Proyecto de Investigación Científico-Tecnológico “Factibilidad y Eficiencia de Aplicaciones en Telemedicina sobre Sistemas Embebidos. Aplicaciones en Arquitecturas Multiprocesador” que se lleva a cabo en la Universidad Nacional Arturo Jauretche (UNAJ).

Por otra parte, se tiene financiamiento en el marco del programa “Universidad, Diseño y Desarrollo Productivo” del Ministerio de Educación a través de los proyectos aprobados en 2014 y 2015.

Parte de las líneas de investigación desarrolladas se encuentran enmarcadas en los convenios de colaboración en Actividades de Investigación firmados por la UNAJ con el Laboratorio CeTAD y el Instituto III – LIDI de la UNLP.

Introducción

En los últimos años, se ha buscado expandir el concepto del procesamiento paralelo multicore hacia plataformas de procesamiento más específico. Para obtener mayor eficiencia, los fabricantes de computadoras de altas prestaciones, han introducido unidades FPGA en su diseño como soporte de procesamiento [1],[2] y [3].

En estudios realizados en laboratorio se llega a la conclusión que las mejoras con estos dispositivos se pueden explorar en los siguientes parámetros:

Comunicación: posibilidad de ensayar diversas topologías conocidas, o explorar nuevas alternativas que podrían optimizar el rendimiento general. Los dispositivos más modernos permiten interfaces de alta velocidad como GbEthernet, o PCI Express. Se pueden implementar protocolos de comunicación flexibles y variar el ancho de bus según la necesidad. Una de las características que se puede utilizar con esta tecnología es la capacidad de reconfiguración parcial de los dispositivos de lógica programable. Esto último permite crear redes virtuales reconfigurables en hardware en tiempo real que sirvan de soporte para enrutado en software [4].

Memoria: cada problema específico requiere de una configuración de memoria determinada. Para ello se cuenta con memoria interna (en general limitada) la cual se podría utilizar para procesos locales y luego se puede interfacear con

distintas tecnologías de memoria existentes. Los dispositivos más modernos tienen incorporados módulos de control para memorias del tipo DDR3. [5]

Software: en este tipo de plataformas es el aspecto más difícil de estandarizar. El software necesariamente es híbrido entre partes de hardware (manejado por drivers) y lenguajes de alto nivel. En estos casos el desafío es encontrar que partes de alto nivel son las que generan mayor “overhead” sobre el procesamiento, identificarlas y luego trasladarlas a su versión de compuertas lógicas. [5]

Si bien el estudio de sistemas paralelos con múltiples procesadores, es una ciencia bien estudiada, el estudio de utilización de múltiples dispositivos reconfigurables no es un terreno completamente explorado [6].

Plataformas FPGAs para procesamiento paralelo

La implementación de paralelismo en plataformas FPGAs consiste en el uso de procesadores embebidos para ejecutar la aplicación, y en la concurrencia que provee la lógica programable para manejar las porciones de código que se pueden ejecutar de dicha forma [7].

La posibilidad de implementar procesadores embebidos en forma rápida [8], junto con la posibilidad de obtener concurrencia mediante la programación de HW permiten combinar el paralelismo con los sistemas multicore en virtud de buscar la máxima eficiencia, lo cual aparece como un gran desafío para lograr la optimización en el procesamiento de imágenes médicas, entre otras aplicaciones. Esto se debe a que cada algoritmo es un caso de estudio, en el que se debe decidir qué parte se implementa de manera concurrente y cuál de forma

secuencial dentro el sistema multi-core. Para la implementación de estos sistemas complejos lo ideal es combinar la programación en VHDL con lenguajes de alto nivel y aplicaciones como el simulink [8]. Con las mejoras constantes que aporta la evolución de la tecnología sobre las FPGAs, se logran diseños de gran magnitud, a tal punto que la tendencia actual es implementar microprocesadores de propósito general y todo el hardware de propósito específico que requiere la aplicación dentro de una FPGA.

Aplicaciones de Transmisión y procesamiento de datos

Con respecto a la transmisión de datos Ethernet por UDP [9], el envío de grandes cantidades de paquetes de datos se realiza detectando si hay una conexión, luego si la hay se define la dirección de destino como broadcast y la mac de origen [10]. Luego, cuando se detecta la conexión se arman los paquetes y se envían a destino desde el Sistema Embebido basado en el LPC1769 [10].

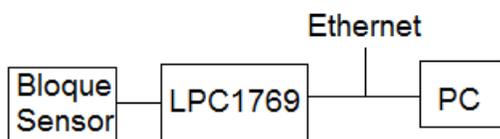


Figura 1. Esquema de comunicación

En la Figura 1 se muestra el esquema de conexionado, donde los paquetes se reciben en una PC y mediante wireshark se pueden determinar tiempos de envío de paquetes y estimar teniendo en cuenta el peor caso, una frecuencia de envío mínima para las aplicaciones de interés.

El método de transmisión es fundamental para el envío de datos en forma masiva, que principalmente consisten en imágenes médicas de alta calidad, hacia un servidor de

almacenamiento exclusivo para telemedicina [11]. Los métodos de computo paralelo y de transmisión de datos para telemedicina siguen la tendencia analizada en [12] y [13].

Previo a ser enviadas, dichas imágenes requieren ser procesadas, mediante la aplicación de diferentes algoritmos de procesamiento de imágenes que tienen características paralelizables y son los que posibilitan evaluar la eficiencia y el rendimiento de las arquitecturas paralelas estudiadas.

Adicionalmente, se realizarán pruebas sobre la arquitectura paralela de señales obtenidas de un sistema MIOC, cuya salida eléctrica es continua y es proporcional a la fase de la señal luminosa que lo atraviesa, este dispositivo permite obtener el ángulo de inclinación en multiplicidad de aplicaciones.

Antecedentes del Grupo de Trabajo

Los antecedentes del grupo de trabajo se parten de la dirección de Tesis de Grado sobre la implementación de web server embebido orientado a telemedicina [14] y de un sistema embebido de voz sobre IP [15]. Posteriormente se han realizado tareas de investigación de procesamiento de imágenes sobre diferentes arquitecturas de procesadores [16], [17] y [18].

Las líneas de investigación se enmarcan fundamentalmente en la búsqueda de técnicas innovadoras de procesamiento paralelo en diferentes arquitecturas que optimicen las prestaciones de los sistemas de telemedicina que puedan ser incorporadas o aplicadas en el Hospital HEC. Adicionalmente, se pretenden encausar estas líneas de investigación en la realización de estudios de postgrado tanto para docentes como para los futuros

profesionales que surjan de la UNAJ. El postulante a director del proyecto participa en proyectos de investigación desde el año 2005 en la Facultad de Ingeniería de la UNLP como colaborador y a partir de 2010 como investigador. Actualmente participa como miembro en los proyectos “Procesamiento Digital de Imágenes Médicas sobre plataformas FPGA”, “Procesamiento de Imágenes en arquitecturas FPGA multiprocesador” y “Envío de datos Ethernet para telemedicina”, pertenecientes a estudios de Maestrías en realización en el Laboratorio CeTAD, en la UNAJ y al convenio UNAJ-LIDI como generador de los lineamientos de investigación propuestos en la UNAJ.

Como resultado del trabajo realizado recientemente en la UNAJ se han publicado artículos; como autor en el 3° Congreso Nacional de Ingeniería Informática y Sistemas de la Información (CoNaIISI-2015), en el VI congreso de Microelectrónica Aplicada (uEA 2015)[18], en el XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015) y en el III Congreso Virtual de Microcontroladores y sus aplicaciones (CVM 2014).

Se destaca la participación de alumnos de la Cátedra Organización y Arquitectura de Computadores, (en la que el director del proyecto actualmente es coordinador), como autores del trabajo presentado en el 3° Congreso de Microcontroladores y sus aplicaciones (CVM-2014) y en el CONAIISI 2015. Por otro lado se debe destacar la participación de los integrantes del proyecto de investigación como autores en los congresos WICC 2015, uEA 2015 y CONAIISI 2015 y LACCEI 2014 [16].

Líneas de Investigación, Desarrollo e Innovación

Temas de Estudio e Investigación

- Implementación de un sistema multicore en Dispositivos Lógicos Programables (FPGA's).
- Búsqueda de eficiencia en el procesamiento de imágenes mediante computo paralelo.
- Estudiar la frecuencia de actualización de datos y eficiencia en la visualización del estado actual de las variables sobre Sistemas basados en Microprocesadores Cortex
- Investigación relacionada con los protocolos TCP y UDP para el envío de paquetes, donde se deben determinar las tasa de transmisión mínimas en condiciones de red normales para asegurar un funcionamiento óptimo del sistema.
- Estudio de las características paralelizables de algoritmos de procesamiento de imágenes en el plano espacial y frecuencial.
- Estudia de las características paralelizables de las señales obtenidas de dispositivos MIOC y su procesamiento.
- Definir parámetros de procesamiento para explotar al máximo las características multicore.
- Estudiar la factibilidad en la transmisión de imágenes, gráficas de datos evolutivos y voz en tiempo real en Sistemas Embebidos .

Resultados y Objetivos

Investigación experimental

- Lograr paralelizar eficientemente determinados algoritmos de

procesamiento de imágenes mediante su implementación sobre plataformas multicore basadas en FPGAs.

- Obtener tiempos de transmisión de datos UDP acorde a la optimización obtenida en el procesamiento paralelo.
- Verificar la eficiencia comparando tiempos de cómputo entre algoritmos de procesamiento de imágenes implementados de forma secuencial y su paralelismo en sistemas multicore.
- Los resultados obtenidos actualmente son satisfactorios, teniendo en cuenta que se ha implementado un sistema embebido en tiempo real que permite transmitir datos de imágenes vía UDP en tiempos que rondan un rango variable dentro de los límites esperados.
- Se ha implementado un sistema multicore sobre una FPGA Virtex V, disponible para realizar las pruebas de procesamiento paralelo y medir eficiencia
- También se han implementado multiplicidad de algoritmos de procesamiento de imágenes médicas de manera secuencial
- Para logra la implementación de un sistema complejo y obtener resultados definitivos se pretende lograr el procesamiento paralelo y la transmisión UDP conjuntamente en tiempo real.

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D se participa en el dictado de asignaturas de la carrera de Ingeniería Informática de la UNAJ.

En esta línea de I/D existe cooperación a nivel nacional. Hay 2 investigadores realizando su Doctorado y 2 realizando una Maestría en temas relacionados con Sistemas Embebidos, Multiprocesadores, Software Embebido y transmisión de datos biomédicos mediante UDP.

Referencias

- [1] Oskar Mencer, Kuen Hung Tsoi, Stephen Craimer, Timothy Todman and Wayne Luk, Ming Yee Wong and Philip Heng Wai Leong ,”CUBE: a 512-FPGA Cluster”, Dept. of Computing, Imperial College London , Dept. of Computer Science and Engineering The Chinese University of Hong Kong . (2009)
- [2] Keith Underwood, “FPGAs vs. CPUs: Trends in Peak Floating-Point Performance”, Sandia National Laboratories . (2011)
- [3] Encinas, Diego, Kunysz, Eduardo, Szymanowski, Alicia, Morales, Martín Daniel, “Performance de arquitecturas multiprocesador: técnicas de simulación y plataformas reconfigurables”, Instituto de Ingeniería, UNAJ, (2014)
- [4] Dong Yin , Deepak Unnikrishnan, Yong Liao, Lixin Gao and Russell Tessier , “Customizing Virtual Networks with Partial FPGA Reconfiguration”, Dept. of Electrical and Computer Engineering University of Massachusetts (2010)
- [5] Eduardo J. Kunysz, José A Rapallini, Jorge Osio, “Sistema de cómputo reconfigurable de alta performance (Proyecto HPRC)”, 3ras Jornadas ITE - 2015 -Facultad de Ingeniería – UNLP
- [6] C. Rupp, The NAPA Adaptive Processing Architecture. IEEE Symposium on FPGAs Custom Computing Machines. 1998.
- [7] Grama A, Gupta A, Karypis G, Kumar V. “Introduction to parallel computing”. Second Edition. Pearson Addison Wesley, 2003.
- [8] "Model-Based Design with Simulink, HDL Coder, and Xilinx System Generator for DSP", Kiran Kintali and Yongfeng Gu, MathWorks, White Paper (http://www.mathworks.com/fpga_design/simulink-with-xilinx-system-generator-for-dsp.html).
- [9] Richard Stevens, “The Protocols TCP/IP Illustrated, Volume 1”, 1993
- [10] Jan Axelson, “Embedded Ethernet a internet complete”, Lakeview Research

LLC, 2003

- [11] Carlos Eduardo Arcila Gomez, “DISEÑO DE UN ENLACE DE TELEMEDICINA PARA EL HOSPITAL UNIVERSITARIO SAN JUAN DE DIOS DEL QUINDÍO”, Tesis de Grado, Universidad de Quindío, 2010
- [12] VÉLEZ Jorge Alberto. Panorama y tendencias de la telemática en salud, hablando de Telemedicina. Sistemas & Telemática, Universidad Icesi de Colombia. p. 3.
- [13] IEEE, A Unified Multimedia Database System to Support Telemedicine. Transactions of Information Technology in Biomedicine, Vol. 2, No. 3, September 1998.
- [14] Lucas Iogna Pratt, “Implementación de un web server embebido”, Tesis de Grado, Director Jorge Osio, Fac. Ing., UNLP, 2013.
- [15] Juan Eduardo Salvatore, “Desarrollo de un Sistema de voz sobre IP”, Tesis de Grado, Fac. Ing., UNLP, 2012.
- [16] Juan Salvatore, Jorge Osio, Martín Morales, “Detección de objetos utilizando el sensor Kinect”, Guayaquil, Ecuador, LACCEI 2014.
- [17] Jorge Osio, Diego Montezanti, Marin Morales, “Análisis de Eficiencia en Sistemas Paralelos”, Ushuahia, Tierra del Fuego, WICC 2014
- [18] Jorge Osio, Juan Eduardo Salvatore, Martín Morales, “sistema embebido para la transmisión de voz vía TCP/IP”, Universidad de La Matanza, Congreso uEA 2015.
- [19] Jorge Osio, Juan Salvatore, Martín Morales, “Análisis de Eficiencia en Arquitecturas Multiprocesador en aplicaciones de Telemedicina”, Universidad Nacional de Salta, Ciudad de Salta, WICC 2015.

Plataforma Middleware para la Gestión Datos de WSN de Manera Interoperable

Diego Alberto Godoy^{1,2,a}, Eduardo O. Sosa^{1,b}, Hernán Bareiro^{1,c}, Juan de Dios Benitez^{1,d}, Edgardo A. Belloni^{1,e}, Rebeca Díaz Redondo^{2,f}

¹Centro de Investigación en Tecnologías de la Información y Comunicaciones (C.I.T.I.C.) - Universidad Gastón Dachary (UGD)

²Universidad de Vigo, Departamento de Ingeniería Telemática. E. E. Telecomunicación.

^adiegodoy@citic.ugd.edu.ar, ^bes@fceqyn.unam.edu.ar, ^chernanbareiro@gmail.com, ^djuan.benitez@citic.ugd.edu.ar, ^eebelloni@ugd.edu.ar, ^frebeca@det.uvigo.es

Resumen

En el presente trabajo se presenta una plataforma middleware para la gestión de datos obtenidos de redes de sensores inalámbricos (WSN) como parte indisoluble de Internet de las Cosas, permitiendo la publicación de resultados en formatos interoperables. El middleware se ha utilizado en varios casos de estudio, como la interacción con la industria del Té y otros en común con la intendencia municipal, facilitando la recolección de residuos, detectando a aquellos que ameriten ser recogidos, en la senda de proyectos convergentes con *Smart Cities*. La plataforma es una línea dentro del proyecto denominado “Diseño de arquitecturas de soporte a la Internet del futuro y Ambientes Inteligentes para su Aplicación en Ciudades Inteligentes – Fase II” de la Universidad Gastón Dachary.

Palabras clave: Simulación, WebML, Redes de Sensores Inalámbricos, IoT

Contexto

El trabajo se enmarca en el proyecto de investigación denominado “Diseño de Arquitecturas de Soporte a la Internet del Futuro y Ambientes Inteligentes” y “Diseño de arquitecturas de soporte a la Internet del futuro y Ambientes Inteligentes para su Aplicación en Ciudades Inteligentes – Fase

II”, acreditados en la Secretaría de Investigación y Desarrollo de la UGD por RR.19/A/12 y RR.18/A/14 respectivamente.

Se relaciona y articula directamente con diversos proyectos de investigación acreditados en la Secretaría de Investigación y Posgrado (SECIP) de la Facultad de Ciencias Exactas, Químicas y Naturales de la UNaM, enfocados en las temáticas “*Internet del Futuro*” e “*Internet de las Cosas*”, entre los que se incluyen: Proyecto 16Q457 “Hacia la Programación de Sensores Inalámbricos en la Forma Web 2.0”; Proyecto 16Q474 “Simulaciones de Sistemas Modernos de Comunicación”; y Proyecto 16Q519 “Ambientes Inteligentes. Una Mirada a Internet del Futuro”.

Referente a este proyecto existen 5 tesinas de grado en curso, 3 tesinas de grado finalizadas, un trabajo final de Especialización finalizado, un trabajo de fin de Máster en curso y una Tesis Doctoral en etapa de redacción final.

Introducción

El aumento de la capacidad de cómputo, la miniaturización de los componentes y el abaratamiento de los dispositivos hacen posible la realización de la Internet de las Cosas (IoT, Internet of Things) [1] y de la Inteligencia Ambiental [2], en la que los objetos cotidianos contarán con sensores conectados a Internet.

El desarrollo de estas tecnologías brinda oportunidades de aprovechamiento económico e industrial, y a la vez representa un enorme desafío tecnológico. Como parte del desarrollo de las tecnologías de la IoT, las Redes de Sensores Inalámbricos (WSN, Wireless Sensor Network) [3] son un área de álgido estudio y desarrollo [4]. Hoy en día existen WSNs con miles de nodos, recolectando datos en infinidad de entornos, para todo tipo de fines y a un costo relativamente razonable.

Objetivos

El objetivo general del proyecto marco es: "Diseñar arquitecturas de soporte a Internet del futuro y Ambientes Inteligentes para su aplicación a Ciudades Inteligentes". Como objetivos específicos se indican: A) Especificar el Estado del Arte de la Internet del de Futuro y los Ambientes y Ciudades Inteligentes. B) Diseñar una arquitectura interoperable para la publicación y aprovechamiento de datos obtenidos desde Redes de Sensores Inalámbricos. C) Construir middlewares que capturen datos de redes de sensores inalámbricos. D) Diseñar arquitecturas de software mashups como soporte de aplicaciones para Ciudades Inteligentes. E) Simular escenarios de Ciudades inteligentes de redes de sensores inalámbricos. F) Desplegar una red inalámbrica para aplicaciones específicas, en dominios como recolección de residuos, sistemas de alertas tempranas, ubicación de ganado, procesos productivos. G) Realizar pruebas en escenarios de Ciudades Inteligentes que involucren luminosidad, humedad, presencia, presión atmosférica, etc.

Construcción de la Plataforma

Para la determinación de requisitos del middleware se han seguidos los lineamientos propuestos por la Union Internacional de Telecomunicaciones (UIT/ITU) en [5]. Algunos requerimientos, siguiendo la clasificación propuesta por la UIT, son de interfaz, en términos de la definición de los mecanismos de interacción con el sistema, y

otros son de funcionalidad [6]. Esto es, están relacionados al funcionamiento que exhibe el middleware exteriormente. Es así que se proponen considerar también algunos requerimientos no funcionales, para asegurar la cohesión y el desacoplamiento.

Así, en este trabajo se propone organizar los requerimientos del middleware en requerimientos generales y en las tres tareas fundamentales del desarrollo: captura de datos desde la red de sensores, almacenamiento de los datos obtenidos, y publicación de los datos en la web. Estas actividades conforman y administran las actividades de gerenciamiento de una WSN. Para ver detalles de este trabajo puede consultarse [7]

En cuanto a los requerimientos generales el middleware debería ser adaptable a la gran diversidad de escenarios posibles de la IoT. Estos escenarios pueden abarcar desde aplicaciones sencillas en las que el sistema se ejecute en una computadora pequeña, hasta grandes aplicaciones en las que deban ejecutarse servidores de altas prestaciones, o incluso en la nube, para ser capaz de cumplir las demandas y requerimientos de aplicaciones de usuario.

Referente a los requerimientos de captura, el middleware debe ser capaz de obtener los datos desde una WSN y acondicionarlos de tal manera que sean convenientemente almacenados. Entonces, es posible observar dos aspectos: la captura de los datos crudos desde la red y la conversión de los mismos a una representación interna adecuada.

Respecto a la captura, se requiere que el diseño del middleware ofrezca una interfaz de conexión a dispositivos que sea independiente del hardware subyacente. Así se garantiza la flexibilidad, escalabilidad y robustez para conectarse a interfaces como USB, Ethernet u otras, según la necesidad utilizando controladores de dispositivo o programas especializados que puedan requerirse para la interacción con los nodos de la WSN. Algunos fabricantes de sensores, como Libelium [8] o Arduino [9] ofrecen programas tipo sniffers para obtener los datos de la red a través de un comando de consola.

Otro aspecto a considerar en la captura es el parsing, que consiste en el análisis sintáctico de los datos crudos, para transformarlos a una representación interna. Este debe soportar diversos tipos de formatos de entrada, no necesariamente estándares, soportando desde formatos de texto simples, hasta los más elaborados, como XML o JSON o incluso contenido binario.

Respecto al almacenamiento, el proceso es responsable de definir la representación interna de los datos, de manera de unificar el formato en el que estos son comunicados a través de las interfaces del sistema. Adicionalmente debe abstraer las cuestiones relacionadas a la persistencia de estos datos y su posterior recuperación. Debe considerarse además que, debido a la naturaleza heterogénea de los sensores en el contexto de la IoT, no es posible definir un tipo de dato estático para representar la información obtenida desde la WSN. Un aspecto importante es la imperiosa necesidad que el módulo de almacenamiento posibilite la integración con programas de tipo ORM/DBAL (Mapeadores Objeto-Relacional/Capas de Abstracción de Base de Datos).

Por último, cuando el módulo de almacenamiento recupere datos deberá hacer uso de un mecanismo de consultas abstracto, que permita realizar consultas sobre los almacenados. Otro requerimiento es que el diseño considere los tipos más comunes de formatos de publicación basados en hipertexto como HTML, JSON y XML; y que asimismo sea extensible a nuevos formatos. Esto resulta central considerando la variedad de consumidores de datos en la capa de servicios propuestos por la UIT [5].

Diseño de la Plataforma Middleware

Siguiendo los requerimientos, se consideraron tres módulos en el diseño, en correspondencia con las tres etapas del ciclo completo de gerenciamiento de los datos de las WSN. El módulo captura se encarga de la obtención de datos desde la WSN incluyendo la interpretación de los mensajes obtenidos de

la WSN, como también de construir con la información obtenida, un objeto de datos que abstrae la estructura de datos del dominio. El módulo de almacenamiento permite la persistencia y consulta de los datos. El módulo Publicador realiza las labores de publicación en la Web. El esquema propuesto se presenta en la Figura 1.

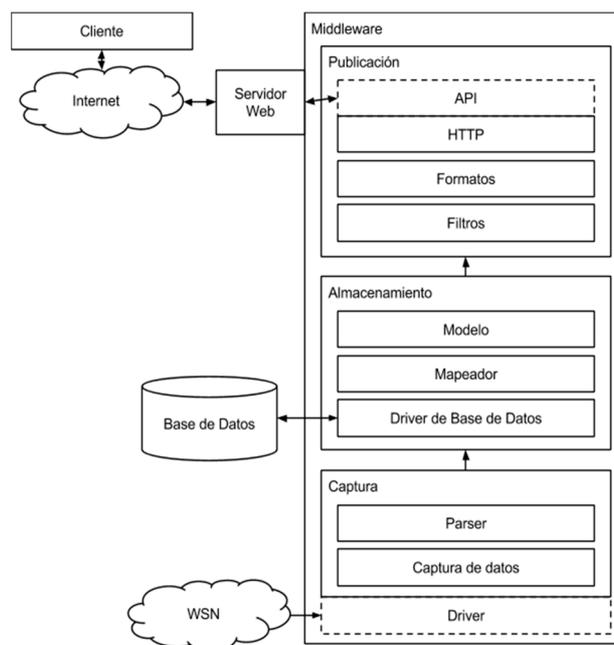


Figura 1 – Esquema propuesto de la solución

El módulo de captura incluye la abstracción de la manera en que se obtienen los datos desde la WSN para luego realizar el análisis sintáctico de las cadenas y las convertirlas a una representación interna. El formato de los datos provenientes de los nodos sensores, los definen los desarrolladores del firmware del nodo. Para cada formato de datos que puedan proveer los nodos, se necesitará extender la clase Parser, para que se pueda reconocer el formato.

El módulo de almacenamiento representa la abstracción de la forma de almacenamiento en tablas físicas e independiza al resto de middleware, del ORM o motor de base de datos.

El módulo de publicación opera sobre HTTP y está basado en REST (*Representational State Transfer*) y permite múltiples formatos como interoperables como

JSON y XML y otros formatos como HTML y CSV.

Testeo de Plataforma

La plataforma middleware, se probó en un escenario concreto tomado del trabajo "Sistema de Monitoreo de la Temperatura y la Humedad para el Proceso de Fermentado en la Industria de Manufactura del Té Negro Basado en una Red de Sensores [10] El trabajo referenciado tenía como objetivo la asistencia en la toma de decisiones respecto al control de los reguladores de temperatura brindando una herramienta de monitoreo a partir de datos obtenidos desde una WSN. El montaje se materializó con nodos ISense [11]: un gateway y dos nodos con módulos climáticos. El sistema cliente consulta al middleware el promedio por minuto de los datos de temperatura y humedad relativa recibidos desde los sensores. Esta consulta se realiza mediante la URL (este ejemplo es para obtener los datos promedio por minuto de temperatura, humedad relativa del sensor 0x1c34):

<http://nodeaddress/middleware/datos?format=json&avg=temp1,rh&avg-interval=minute&sensorid=0x1c34>

En la figura 2 se muestra la interfaz de usuario del sistema y el formato de representación interna de los datos obtenidos de los sensores.

Este ejemplo muestra la utilidad práctica de contar con un middleware, permitiendo reducir el tiempo de desarrollo, gestionando la comunicación con la WSN subyacente, y realizando la persistencia de datos de manera transparente al sistema cliente. De igual manera ofrece ayuda al momento de obtener los datos, proveyendo información de resumen. Paralelamente, el diseño interno del middleware demuestra en este caso su extensibilidad (se agregó el dato de ID de sensor), permitiendo la adición de un nuevo parámetro sin necesidad de realizar modificaciones de consideración. El middleware permite así independizarse de aplicaciones de terceros que no abarcan el

ciclo completo de gestión de datos de WSN. El middleware reemplazó al desarrollo que se había realizado específicamente para la medición del de la temperatura del Té, posibilitando extender la funcionalidad en el futuro para este dominio.

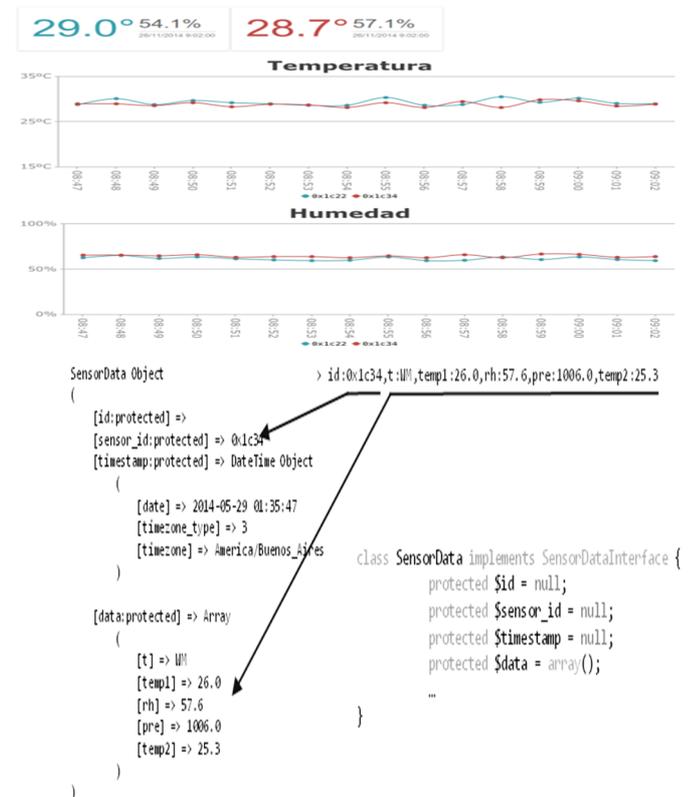


Figura 2: Interfaz de usuario y formato de cadena, Objetos Json y clase SensorData

Resultados

Se ha presentado una plataforma útil a los desarrolladores de aplicaciones orientadas de IoT de las peculiaridades de las distintas redes de sensores inalámbricas, formatos de datos, motores de gestión de persistencia bases datos y formatos de datos interoperables. Esta abstracción se presenta en formato de framework, el cual se puede extender según la aplicación y la WSN subyacente. El middleware permite capturar, almacenar y publicar datos en formatos interoperables. Esto constituyen una alternativa de soporte a aplicaciones Mashups como las estudiadas en

[12] aunque multiplataforma y extensible. Se han realizado pruebas del ciclo completo del middleware con resultados satisfactorios. El prototipo se ha utilizado en otros proyectos en ámbitos de ciudades inteligentes [13] y de interacción con industrias de la zona [10]; reemplazando los desarrollos específicos existentes.

Formación de Recursos Humanos

El equipo de trabajo se encuentra formado por una Doctora en Ingeniería Telemática, un Doctor en Ciencias Informáticas, un Doctorando en Tecnologías de la Información y Comunicaciones, Magister y Especialista en Ingeniería de Software, un Maestrando de Ingeniería de la Web, un Maestrando de Redes de Datos, tres auxiliares de investigación graduado, un auxiliar de investigación graduado residente en la ciudad de Córdoba Capital, y ocho auxiliares de investigación en período de realización de trabajos de grado. El número de tesinas de grado en curso con proyecto aprobado es de cinco y el número de trabajos de especialidad finalizado es tres dentro de la línea de investigación. Los proyectos de grado se titulan “Diseño de un prototipo para monitoreo eficiente de iluminación basado en WSN utilizando HTML5”, “Contribución a la Gestión de Residuos Domiciliarios como una Aplicación en Ciudades Inteligentes” y “Análisis y comparación de modelos de propagación para optimizar la localización geográfica de Ganado”.

Bibliografía

- [1] K. Ashton, "That 'Internet of Things' Thing", 2009 (rev. 2011).
- [2] Ahola J., *Ambient Intelligence*, 2001.
- [3] W. Dargie and C. Poellabauer, *Fundamentals of Wireless Sensor Networks - Theory and Practice*. Reino Unido: Wiley, West Sussex, 2010.
- [4] Eduardo O. Sosa, *Contribuciones al establecimiento de una red global de Sensores Inalámbricos*. Tesis Doctoral, Junio 17, 2011.
- [5] International Telecommunication Union ITU, *Service description and requirements for ubiquitous sensor network middleware*.: ITU-T Recommendation F.744, 2009.
- [6] IEEE, *IEEE Recommended Practice for Software Requirements Specifications*. New York, 1998.
- [7] Diego Alberto Godoy, Eduardo Omar Sosa, and Rebeca Díaz Redondo, "Internet De Las Cosas: Middleware De Gestion De Datos De WSN," in *CONFERENCIA WWW/INTERNET 2015 e COMPUTAÇÃO APLICADA 2015 FLORIANÓPOLIS*, Florianópolis, 2015.
- [8] Libelium Comunicaciones Distribuidas S.L., *Waspmote Technical Guide*., 2014.
- [9] Arduino. (2014) USB Serial Light Adapter. [Online]. <http://bit.ly/20MsGMg>
- [10] P. Quiñones, D. Godoy, and E. Sosa, "Redes Inalámbricas de Sensores: Una Experiencia En La Industria Del Té," in *42º Jornadas Argentina de Informática*, Córdoba, 2013.
- [11] Coalesense. (2015) iSense Wireless Sensor Network Software. [Online]. <http://bit.ly/1WMMiR0>
- [12] Diego Alberto Godoy, *Plataformas para la creación de mashups sensibles al contexto en entornos de inteligencia ambiental*. Trabajo Final de Especialidad en Ingeniería de Software, 2013, Universidad Nacional de Plata.
- [13] E. Sosa et al., "Internet del futuro y ciudades inteligentes," , Paraná, 2013.

Desarrollo de la comunicación Peer to Peer para Sistemas de Tiempo Real basado en IPV6

Isabel B. Marko, Gonzalo Vilanova,

Jonathan Kernbeis, Mariano Sabala

Escuela de Sistemas

Universidad John F. Kennedy.

Bartolomé Mitre 1411 – CP (1037) – CABA.

5236-1224/1216/1242

imarko@kennedy.edu.ar, gvilanova@kennedy.edu.ar, jkernbeis@kennedy.edu.ar,
msabala@kennedy.edu.ar,

Resumen

El protocolo IPV6 no solo representa un aumento fabuloso en la cantidad de nodos disponibles, unívocamente identificados en el mundo, sino que además incluye una serie de mecanismos que lo hacen altamente preferible al IPV4. Por ello se analizará la utilización del protocolo de red IPv6 a nivel aplicación, viendo como las aplicaciones nativas, pueden verse beneficiadas en la implementación de redes P2P de tiempo real.

Palabras clave: IPv6, Peer to Peer, Tiempo Real.

Contexto

El proyecto es financiado y desarrollado en la Universidad Argentina John F. Kennedy (UK), y forma parte de la línea de proyectos de investigación de la Escuela de Sistemas.

Introducción

Las redes P2P están compuestas por esquemas donde cada nodo es equivalente a los otros, o sea todos los nodos pueden funcionar como clientes y como servidores, la información puede estar distribuida y replicada en una gran cantidad de nodos, o simplemente estar solo incluida en los nodos enlazados (Figura 1).

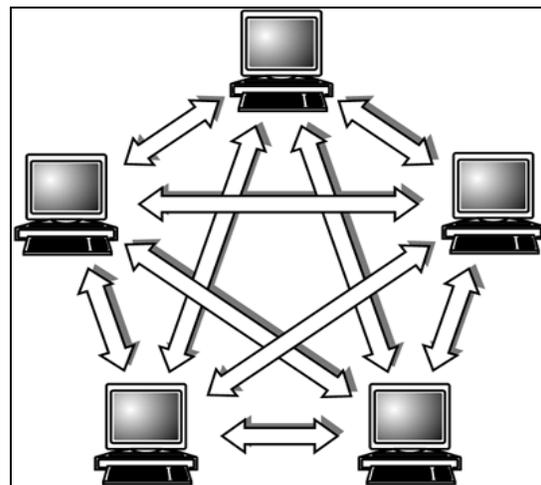


Figura 1. Red P2P

Las comunicaciones Peer to Peer (P2P) de tiempo real son el cimiento de algunas de las aplicaciones más populares en Internet.

Si bien IPv6 es un protocolo de red sus nuevas características permiten beneficiar a diversos tipos de aplicaciones, entre ellas las que:

- Manejan grandes volúmenes de datos;
- Requieren de streaming de audio y video;
- Almacenan información en la nube;
- Requieren una calidad de servicio mínima y un control en las metas temporales de la comunicación. Fundamental en sistemas y aplicaciones de tiempo real, tales como la telefonía ;
- Brindan soporte a la Telemedicina;
- Permiten el aprendizaje a Distancia (e-learning);
- Utilizan redes de Sensores

Por lo dicho previamente IPv6 tiene diversas ventajas que pueden ser aprovechadas al momento de desarrollar aplicaciones Peer to Peer de tiempo real, nativas para este protocolo.

Hay unas 25 mil millones de “cosas” conectadas a Internet y se estima que en el 2020 serán 50 mil millones los objetos conectados. Todas esas “cosas” transmitirán información.

Se deben desarrollar aplicaciones Peer to Peer de manera confiable y cumpliendo con las metas temporales que en un sistema de tiempo real se exigen, así como con los requisitos de seguridad y confidencialidad en la información intercambiada. Actualmente coexisten en funcionamiento, el protocolo de internet IPV4 e IPV6, dependiendo en gran

medida del lugar que protocolo se esté utilizando.

Redes de este tipo son las que utilizan por ejemplo en la tecnología BitTorrent [4]. En teoría cada nodo puede ser un cliente y un servidor dentro de la red. La escalabilidad de sistema es potencialmente muy flexible. P2P requiere de ciertas características que el protocolo IPv4 en su forma nativa estaban levemente o para nada soportadas, y cuya implementación posterior era complicada debido fundamentalmente a la limitada capacidad de direccionamiento relativa del IPv4, que limita la unicidad de direcciones IP en todo el dominio de Internet.

Por ejemplo en una red P2P de tiempo real, cada nodo debe ser capaz de:

- Descubrir otros clientes
- Conectarse con otros clientes
- Comunicarse con otros clientes
- Garantizar las metas temporales
- Atender los temas de seguridad en caso de información sensible.

IPv6 surgió sobre todo para solucionar el problema de agotamiento de direcciones IPv4.

Líneas de Investigación, Desarrollo e Innovación

El presente proyecto se encuentra en etapa de desarrollo, siendo los puntos a tratar los siguientes:

- Estudiar en profundidad las ventajas que el protocolo IPV6 ofrece en su utilización en sistemas de tiempo real con comunicación Peer to Peer.
- Desarrollar una aplicación Peer to Peer en IPV6 que permita la

transferencia de datos en tiempo real entre 2 móviles conectados a una red 4G o 3G.

- Extender la aplicación del sistema a diversas áreas debido a la diversidad de datos (datos genéricos)

Resultados y Objetivos

A partir de los resultados obtenidos, los futuros objetivos son:

- Utilizar el protocolo IPV6 en la implementación de redes Peer to Peer para sistemas de tiempo real, lo que facilitará de gran manera el desarrollo de aplicaciones, utilizando hardware y redes disponibles que soporten el protocolo.
- Promover una aplicación para teléfonos inteligentes, que nos permita el intercambio de datos de forma genérica y cumpliendo con las pautas de los sistemas de tiempo real aprovechando las redes 4G y 3G que soportan IPV6

Formación de Recursos Humanos

En esta línea de investigación se encuentran afectados docentes de la universidad pertenecientes a las áreas de informática y electrónica. Así mismo alumnos pertenecientes a la carrera de Licenciatura en Sistemas, donde desarrollan sus trabajos de fin de carrera.

Referencias

- MSDN Library
[http://msdn.microsoft.com/es-es/library/bb902818\(v=SQL.105\).aspx](http://msdn.microsoft.com/es-es/library/bb902818(v=SQL.105).aspx)
- Microsoft Press. Peer - to - Peer Networking.
- http://blogs.wrox.com/files/2013/05/502259_c45_p2_ed.pdf
- Windows – Dev Center - Desktop
- [http://msdn.microsoft.com/en-us/library/windows/desktop/bb153250\(v=vs.85\).aspx](http://msdn.microsoft.com/en-us/library/windows/desktop/bb153250(v=vs.85).aspx)
- BitTorrent is a P2P file sharing client or torrent client, letting users to download and upload large files
- <http://www.bittorrent.com/intl/es/>
- FEYRER Hubert, O’ Reilly. “The future of the Internet”. 2001.
- http://onlamp.com/pub/a/onlamp/2001/05/24/ipv6_tutorial.html
- D. Giulianelli, R Rodríguez, P Vera, M Cornejo. “Desarrollo de Aplicaciones Nativas para IPv6”. 2013.
- <http://sedici.unlp.edu.ar/bitstream/handle/10915/27093/Desarrollo+de+Apl+icaciones+nativas+para+IPV6.pdf?sequence=1>
- L. Atzori, A. Iera, G. Morabito, “The Internet of Things: A Survey”, J. Computer Networks, Vol.54, no.15, Oct. 2010, pp 2787-2805G.
- A.P. Castellani, et al, Architecture and protocols for the Internet of things: A case study”, Proc. 1st IEEE Int. Workshop on Web of Things (WoT 2010), IEEE PERCOM, 2010, pp 678-683
- D. Guinard, V. Trifa, E. Wilde, “Architecting a mash able open world wide web of things”, TRCS-663 ETH Zurich, 2010
www.vs.inf.ethz.ch/publ/paprs/WoT.pdf
- Internet of Things Consortium.
<http://iofthings.org/>

Explorando posibles mejoras de protocolo TCP en redes móviles

Diego R. Rodríguez Herlein Carlos A. Talay

Campus Universitario – oficina 18 - Dpto. Ciencias Exactas e Informática UARG / UNPA
diegorh@gmail.com carlostalay@yahoo.com.ar

Luis Marrone

L.I.N.T.I. – Universidad Nacional de La Plata
Calle 50 y 115 – 1er. Piso – Edificio Bosque Oeste
lmarrone@info.unlp.edu.ar

Resumen

En este trabajo se intenta realizar una mínima recopilación de los esfuerzos que se han realizado en adaptar al protocolo TCP (Transmission Control Protocol) a redes de datos en donde las tasas de error son mucho mayores que las tasas para el cual fue pensado originalmente este protocolo, debido a la congestión de datos. Así mismo intentaremos analizar que posibles mejoras pueden realizarse en la aplicación del protocolo TCP en redes móviles, en donde un enlace wireless interviene en uno o más de los trayectos de los datos e introduce dos fenómenos típicos de este enlace como son, la pérdida de paquetes por diversas causas o directamente el desacople de uno de los nodos.

Palabras clave: TCP, performance, wireless, control congestión

Contexto

Este estudio está enmarcado en el PI 29/A358-1 “Análisis de performance del protocolo TCP utilizado en redes móviles” radicado de la UNPA-UARG, que se encuentra actualmente en etapa inicial de desarrollo. El proyecto está compuesto mayoritariamente de docentes de la

UNPA-UARG y coordinado por su Director el Sr. Luis Marrone perteneciente a la UNLP. Este proyecto se financia íntegramente con fondos destinados a proyecto de investigación de la UNPA-UARG.

Introducción

TCP [1] es un protocolo orientado a la conexión, seguro y el más utilizado en la capa de transporte. Soporta la mayoría de las aplicaciones más populares de internet. Es fiable y se adapta adecuadamente a las redes cableadas, de nodos estáticos, que regula el número de paquetes enviados mediante el método de ventanas deslizantes [2], el cual lleva el control de paquetes enviados y controla su recepción mediante el envío de un mensaje de ACK. De esta manera, TCP se adapta a los problemas de las redes cableadas, debidos al retraso y pérdida de paquetes por congestión [3,4]. Los algoritmos de control de congestión, tratan de determinar dinámicamente el ancho de banda y la latencia de la red, modificando el envío de paquetes y su rendimiento, a los cambios de tráfico para evitar el colapso de la subred.

Este protocolo fue desarrollado y posteriormente optimizado para redes cableadas, lo que implica que la pérdida de paquetes se debe casi con exclusividad a la congestión en la red, suponiendo una tasa de error de bits, en tránsito, prácticamente inexistente. TCP realiza el control de congestión regulando el tráfico que inyecta a la red. Para esto, debe conocer cuando es posible enviar más tráfico y cuando se alcanzó su capacidad máxima.

El control de congestión comienza con el algoritmo de arranque lento (slow start), en donde el TCP emisor comienza con una Ventana de Congestión (cwnd) con valor 1 [5]. Por cada ACK recibido incrementa el valor de dicha ventana en forma exponencial hasta un valor umbral (ssthresh), en donde comienza el algoritmo evitar congestión (congestion avoidance) [6]. En este punto se comienza a incrementar cwnd en forma lineal hasta el tamaño máximo de la ventana del receptor. Asimismo, TCP mide el tiempo que tardan en llegar los ACK, manteniendo un promedio de este retraso y la desviación esperada de ese promedio. Entonces, si el retraso del ACK de un paquete es mayor que el promedio incrementado en 4 veces el desvío calculado o bien recibe un ACK duplicado, el TCP asume que el paquete se perdió y lo retransmite. Es en este momento que se activa el algoritmo de Retransmisión Rápida (fast retransmit).

TCP reacciona a cualquier pérdida de paquetes reduciendo el umbral (ssthresh) a la mitad o a 2 (lo que sea

mayor), restableciendo el valor de la ventana de congestión a 1, activando el algoritmo de arranque lento y, estableciendo el temporizador de retransmisión al intervalo de backoff de acuerdo al algoritmo de Karn [7].

Dentro de las implementaciones de TCP existentes, se destacan [8]:

TCP Tahoe: Posee los algoritmos de slow start y congestion avoidance y posteriormente se le agrego el algoritmo de fast retransmit. La principal desventaja de esta implementación es que cada vez que se pierde un segmento, comienza nuevamente con el algoritmo de slow start, lo que demora demasiado el retorno al valor de las tasas anteriores.

TCP Reno: Cuando se detecta congestión, el TCP Tahoe reduce la ventana de congestión a un segmento lo que dispara el algoritmo de slow start, mientras que TCP Reno dispara el algoritmo de Fast Recovery, reenviando el paquete perdido y reduce el umbral a la mitad, evitando disparar el algoritmo de slow start. Cuando recibe un nuevo ACK, sale de fast recovery y dispara el algoritmo de congestión avoidance. Esto permite que se recupere más rápido de la congestión.

TCP New Reno: La optimización en esta versión de TCP mejora la pérdida individual y aislada de segmentos pero no mejora las pérdidas en segmentos sucesivos. La diferencia con el TCP Reno es que cuando se está en la etapa de fast recovery, el arribo de uno nuevo ACK no hace que se salga de esta, hasta que todos

los segmentos pendientes de confirmar antes del disparo del algoritmo tengan su acuse de recibo. De esta forma un ACK parcial implica que el segmento siguiente se perdió y debe ser retransmitido.

TCP Selective ACK (SACK):

Implementa un mecanismo por el cual el emisor recibe información de los segmentos recibidos e, implícitamente, los perdidos en el receptor. Para esto se implementa un campo opcional SACK en el encabezado de TCP que se envía cuando se reciben segmentos fuera de orden. Contiene la información de los bloques recibidos y sus números de secuencia. Esta versión de TCP utiliza un control de congestión básico y solo utiliza el temporizador de retransmisión como última opción de recuperación. En forma análoga a TCP Reno, TCP SACK dispara el algoritmo de fast recovery al recibir ACK duplicados. Cuando el emisor recibe un ACK duplicado con la opción SACK, almacena los ACK especificados en SACK, permitiendo retransmitir solamente los paquetes implícitamente perdidos en el receptor. De la misma forma que TCP New Reno, sale de la etapa de fast recovery cuando todos los paquetes retransmitidos tienen su acuse de recibo. La ventaja es que, de esta forma, mejora el manejo de la pérdida de segmentos en la misma ventana de forma más eficiente.

TCP Vegas: TCP Vegas posee grandes modificaciones con respecto a las versiones anteriores de TCP. Hasta ahora la ventana de congestión crecía hasta que ocurría una pérdida de paquete y cuando

esto pasaba, el throughput de la conexión se degrada. El TCP New Reno y anteriores asumen que hay congestión solamente cuando detectan estas pérdidas de paquetes. La idea del TCP Vegas es controlar y mantener el tamaño adecuado de la ventana de manera que no ocurra la pérdida de paquetes y se evite que se degrade el throughput. El TCP Vegas administra el tamaño de la ventana observando RTT (round-trip-times) de los paquetes que el TCP emisor envió con anterioridad. El RTT es el tiempo desde que el emisor envía el paquete hasta que arriba el ACK correspondiente. Si el RTT es mayor, TCP Vegas reconoce que la red comienza a estar congestionada y achica la ventana. Por el contrario, si los valores de RTT son más pequeños, el TCP emisor determina que la red no está en congestión e incrementa el tamaño. Otra de las modificaciones importantes del TCP Vegas es en el algoritmo de control de congestión, en la fase de slow start. La tasa de crecimiento del tamaño de la ventana, en esta fase, es menor que en TCP Tahoe y TCP Reno que aumenta su tamaño cada vez que reciben un ACK. Contrariamente, si se recibe un ACK y el RTT es idéntico al promedio, la ventana se mantiene sin cambios.

Con el crecimiento en la demanda de movilidad, el protocolo TCP se comenzó a utilizar sobre enlaces inalámbricos. Los enlaces inalámbricos tienen características de transmisión muy diferentes a los enlaces tradicionales cableados. Estas características propias tienen su origen tanto en la naturaleza del medio físico como en los efectos debidos

a movilidad de los host. Un host es móvil cuando puede moverse libremente alrededor de un área y permite a los usuarios acceder a los datos en cualquier momento y en cualquier lugar, sin diferenciar la operación ni la performance entre la movilidad y un host fijo. De esta manera, los retrasos y las pérdidas de paquetes en las redes inalámbricas, ya no se deben exclusivamente a la congestión, sino también a la pérdida de paquetes por daño en tránsito y la desconexión temporal y frecuente de los nodos, lo que degrada el rendimiento del TCP. El comportamiento de los enlaces inalámbricos está limitado por las frecuentes ráfagas de error causadas por los desvanecimientos, atenuaciones e interferencias. Durante los periodos de ráfaga, todos los intentos de transmisión de datos resultan fallidos con muy alta probabilidad. Durante las desconexiones frecuentes, producto de la misma movilidad, se produce un intervalo de tiempo en el cual no se puede enviar ni recibir paquetes lo que produce aumento en la latencia. Las pérdidas en enlaces inalámbricos se deben principalmente a la pérdida de paquetes en tránsito y no a la congestión. Dado que el protocolo TCP supone que las pérdidas se deben a la congestión, en cada uno de estos casos activara el algoritmo de control de congestión pudiendo ocasionar un comportamiento muy poco eficiente del protocolo en ese entorno [9].

Uno de los esquemas interesantes para diferenciar la causa de la pérdida de los paquetes, se basa en la coherencia de la congestión entre paquetes consecutivos,

llamado Congestion Coherence [10,11]. Se basa en la idea que la congestión no ocurre repentinamente, es decir, ya hay congestión antes de que se torne severa y se descarte un paquete y tampoco desaparece repentinamente luego de que un segmento es desechado.

Otro enfoque proactivo, pensado para evitar la congestión, es el algoritmo denominado RED (Random Early Detection). Mientras que los algoritmos tradicionales descartan paquetes a partir de un buffer lleno, RED utiliza un enfoque probabilístico para el descarte de paquetes de ese buffer. Este procedimiento de descarte de paquetes se realiza antes de producirse un período de alta congestión. El cálculo lo realiza en base al tamaño promedio de la cola, el tamaño del paquete, la cantidad de paquetes desde el último descarte y valores umbrales de referencia máximo y mínimo.

Líneas de Investigación, Desarrollo e Innovación

El protocolo TCP es claramente el más utilizado en sistemas de transmisión sobre redes de datos. Sin embargo la evolución de los sistemas de comunicaciones ha generado una serie de variantes en las cuales se mezclan sistemas cableados con wireless, o sencillamente redes wireless con nodos de mucha movilidad [12]. Esta combinación a dado origen a nuevos problemas basados fundamentalmente en la enorme diferencia de tasa de errores que normalmente estos dos sistemas tienen, más los problemas típicos de desconexión de las redes wireless. Estos problemas han puesto en un verdadero aprieto a la versión original del protocolo TCP, el cual ha sido modificado, generado

distintas versiones que tienden a dotar de estabilidad a este protocolo bajo las distintas variantes de conexionado. En este contexto, diferentes grupos de investigación siguen proponiendo distintas modificaciones o técnicas que tienen a mejorar las condiciones de comunicación. Sin embargo es muy dificultoso llegar a una solución general debido a la heterogeneidad de las topologías, las variantes de los sistemas de comunicación y particularmente los diferentes requerimientos que existen sobre los distintos tipos de redes de datos.

Resultados y Objetivos

Si bien el presente proyecto aún no ha registrado un grado de avance que permita mostrar resultados, el análisis de las soluciones existentes posibilita inferir que las constantes modificaciones del protocolo TCP para mejorar su rendimiento en medios wireless, plantea la seria posibilidad de desarrollar un nuevo protocolo que ataque de cero la problemática. En tal sentido las soluciones existentes del protocolo TCP que contempla una reacción proactiva, parece ser la más adecuada. Por ello las soluciones propuestas por protocolos como el TCP Vegas y RED (Random Early Detection), que reaccionan variando la ventana de congestión en función de prever el estado de congestión de la red analizando el RTT de cada paquete o variando el descarte de tramas en función del estado del buffer de retransmisión, parecen ser un buen punto de partida en desarrollos que se adapten a las exigentes demandas de las redes wireless.

Formación de Recursos Humanos

El grupo de investigación conformado se caracteriza una constitución heterogénea de profesionales vinculados a

la informática. Entre ellos podemos enumerar una Ingeniera en Sistemas de Información, Licenciados en Informática, Ingeniero en Telecomunicaciones, Ingeniero en Electrónica y un Ingeniero Electricista con un máster en Sistemas y Redes de Comunicaciones. También integra el grupo un alumno de la carrera Lic. en Sistemas de la UNPA-UARG. En el transcurso del proyecto se tiene como objetivo consolidar la formación en investigación de los integrantes de menos antecedentes en proyectos y también está contemplado que uno de los integrantes complete su trabajo final de maestría.

Referencias

- [1] Postel J., "RFC 793: Transmission Control Protocol", September 1981.
- [2] M. Handley, J. Padhye and S. Floyd, "TCP Congestion Window Validation", RFC 2861, June 2000.
- [3] V. Jacobson, "Congestion Avoidance and Control," ACM SIGCOMM Computer Communication Review, Vol. 25, No. 1, 1995, pp. 157-187.
- [4] Allman M., Paxson V. and Blanton E., "TCP Congestion Control", IETF, Standards Track RFC 5681, Sept. 2009.
- [5] Stevens, W., "TCP slow start, congestion avoidance, fast retransmit, and fast re-covery algorithms". RFC 2001, 1997.
- [6] Jacobson V., "Congestion Avoidance and Control", ACM SIGCOMM Computer Communication Review, 25(1), 1995, pp. 157-187.
- [7] Karn, H. & Craig P., "Improving round-trip time estimates in reliable transport protocols", ACM Transactions on Computer Systems (TOCS) 9 (4), 364-373, 1991.
- [8] Teja F. R., Vidal, L., Alves, L., "TCP sobre enlaces wireless – Problemas y algunas posibles soluciones existentes", Curso de posgrado y actualización, Instituto de Ingeniería Eléctrica, Facultad de la República, marzo 2004.
- [9] Elaarag, H., "Improving TCP Performance over Mobile Networks", ACM Computing Surveys, Vol. 34, No. 3, September 2002.
- [10] C. Liu, Raj Jain, "Approaches of Wireless TCP Enhancement and A New Proposal Based on Congestion Coherence", 36th Hawaii International Conference on System Sciences. 2002.
- [11] C. Liu, Raj Jain, "Congestion Coherence: A Local Enhancement for TCP over Wireless", OSU Technical Report, November 2004.
- [12] Ho, L., "An investigation of improving the traditional TCP over MANETs". Thesis Master of Computer and Information Sciences, Aukland University of Technology, 2012.

La conectividad en las zonas rurales

Antonio Castro Lechtaler^{1, 2 y 3}; Alejandro Arroyo Arzubi¹; Fernanda Carmona³;
Antonio Foti⁴; Rubén Fusario¹; Alejandro Oliveros⁵.

¹Laboratorio de Investigación en Técnicas Criptográficas y Seguridad Teleinformática. Escuela Superior Técnica, Facultad de Ingeniería. Instituto Universitario del Ejército, C1426; ²CISTIC/FCE - Universidad de Buenos Aires, C1120; ³Universidad Nacional de Chilecito, Chilecito, Provincia de La Rioja, F5360; ⁴Universidad Nacional de 3 de Febrero, Sáenz Peña, Provincia de Buenos Aires, B1674 y ⁵Universidad Nacional del Oeste, San Antonio de Padua, Provincia de Buenos Aires, B1718.

acastro@est.iue.edu.ar; aarroyo_arzubi@hotmail.com; fcarmona69@gmail.com;
foti.antonio@gmail.com; rfusario@gmail.com; aoliveros@untref.edu.ar.

1. Resumen.

En la Argentina, como en otros países en vías de desarrollo, muchas regiones rurales con baja densidad poblacional carecen de servicios de comunicaciones por falta de interés de las empresas públicas de telecomunicaciones en brindar estas prestaciones. Esta situación dificulta a la población rural la educación, la actividad económica, la atención de la salud en casos de urgencia, y provoca todo tipo de problemas a las comunidades que tienen estas carencias. El presente grupo de investigación organizado en red de varias universidades tiene por objetivo buscar una solución a la problemática de la falta de conectividad en las zonas rurales con el objeto de que se puedan brindar servicios de voz y datos de banda ancha con acceso a la Red Internet.

La idea central de la investigación es buscar distintas alternativas que seguramente diferirán de las utilizadas en los países centrales pero que puede constituir una solución a este problema. En el caso particular de Argentina las distancias son condicionantes y la densidad poblacional es sustancialmente muy inferior al que se puede encontrar en aquellos países.

Es por ello que parecería que las técnicas digitales inalámbricas *wireless technolo-*

gies -en especial aquellas de largo alcance tales como: microondas, 802.11, WiMax, CDMA450, 802.22 y otras similares- podrían dar solución al problema planteado.

Palabras Clave:

PLC, CSMA/CA, WLAN, 802.11, 802.22

2. Contexto.

En Grupo de Investigación trabajando en Red de Universidades viene trabajando en analizar los problemas que se presentan en las comunidades rurales por carecer de conectividad y por tanto de comunicaciones, por ser estas áreas no rentables para las empresas proveedoras de servicios de telecomunicaciones, y simultáneamente buscando soluciones técnicas, a costos razonables, estos problemas.

El Grupo tiene su sede en la Escuela Superior Técnica "Gral. Div. Manuel N. Savio" (EST), Facultad del Ejército - Universidad Nacional de la Defensa (UNDEF) y participan además, la Universidad de Buenos Aires y las Universidades Nacionales de Chilecito, de Tres de Febrero y del Oeste.

Los trabajos han incluido pruebas de campo en base a subsidios obtenidos por parte de la Agencia Nacional de Promoción Científica y Tecnológica [1] y fon-

dos aportados por las universidades participantes.

Actualmente luego de varias pruebas efectuadas utilizando diversas tecnologías existentes en el mercado, se está trabajando sobre la base de analizar y probar las posibilidades que brinda en lo técnico los equipos que responden a la Recomendación 802.22 de la IEEE; y desde el punto de vista de su posible implementación, las instalaciones existentes que se utilizan en los predios que posee el Sistema Argentino de Televisión Digital de la Empresa ARSAT.

CITEDEF y el Consejo Profesional de Ingeniería de Telecomunicaciones, Electrónica y Computación (COPITEC) dieron su aval para este proyecto, dado su interés en ser aplicado en beneficio de los productores rurales, sus propios sistemas y redes.

3. Introducción.

El problema de las comunicaciones rurales ha movilizado a distintos grupos de investigación y empresas de telecomunicaciones a buscar soluciones a este tipo de problemas, por cuanto estas zonas geográficas proporcionan cantidades significativas de productos alimenticios en sus diferentes etapas de fabricación y constituyen una trascendente fuente de productos básicos de exportación e ingresos de divisas. Para muchos países participan generando un porcentaje significativo del producto bruto interno de ellos.

La serie de Recomendaciones 802.XX incluye un conjunto de normas que regulan el funcionamiento de las comunicaciones inalámbricas.

Luego de evaluar con resultados poco satisfactorios los equipos que utilizaban la norma 802.11 fueron apareciendo distintas tecnologías que permitieron ampliar

esta Serie merced al trabajo de distintos grupos de investigación [2], [3], [4], [5], [6].

Estos trabajos culminaron el 1 de julio de 2011 cuando finalizó el proceso de aprobación, del estándar “IEEE 802.22 - “IEEE 802.22: Cognitive Wireless Regional Area Network - Medium Access Control (MAC) and Physical Layer (PHY). Specifications: Policies and Procedures for Operation in the TV Band¹”.

La misma fue aprobada con el apoyo del Comité LAN/MAN² de la IEEE [7].

Este nuevo estándar proporciona una opción que permite establecer enlaces inalámbricos full dúplex a distancias de entre 30 a 70 km entre antenas, utilizando frecuencias no restringidas por las regulaciones gubernamentales.

La norma que pertenece a la serie 802.XX³ tiene por objeto establecer los criterios para el despliegue de múltiples productos interoperables de la misma, ofreciendo acceso a la banda ancha fija en diversas áreas geográficas, incluyendo especialmente los de baja densidad de población en las zonas rurales, y evitar la interferencia a los servicios que trabajan en la televisión de radiodifusión.

La misma es conocida actualmente como Red Inalámbrica de Área Regional y está pensada para operar principalmente como una forma de poder acceder a servicios de banda ancha a redes privadas de datos ubicadas en Zonas Rurales.

¹ “IEEE 802.22 - Red de Área Regional Cognitiva Inalámbrica de Control de Acceso al Medio (MAC) y la Capa Física (PHY). Especificaciones, Políticas y Procedimientos para la Operación en las Bandas de Televisión”.

² LAN: Local Area Network; MAN: Metropolitan Area Network.

³ Redes Inalámbricas.

4. Características particulares de la 802.22 - IEEE.

Las características técnicas de la norma, además de proporcionar una solución al problema que nos ocupa, permiten resolver dos problemas que afectan seriamente el uso del espectro de frecuencias. Uno es el problema de los llamados White Spaces [8, 9], y otro, el de las interferencias entre canales, ambos cada día más frecuentes debido al uso intensivo de las comunicaciones inalámbricas para todo tipo de servicios de comunicaciones.

Dichas características son desde el punto de vista de los sistemas de información muy interesantes, pues combinan un problema de comunicaciones, como es el uso adecuado del espectro, con el desarrollo de software especial que permite emplear las técnicas denominadas “radio cognitivas”.

Las mismas moderan la interferencia que pueden ocasionar otros operadores existentes que trabajan en las mismas frecuencias y otorgan capacidad de geolocalización. Esto se ejecuta a través del acceso a una base de datos de los servicios establecidos, y así para detectar la presencia de otros servicios a través de la tecnología de espectro de detección, Para ello la norma señalada, se utiliza en combinación con otra conocida como WRAN o por su denominación IEEE 802.22.1.

Los sistemas WRAN se sirven de canales que van desde 54 a 862 MHz en las bandas de VHF y UHF. El uso de las tecnologías de radio cognitiva permite utilizar los espacios ubicados entre dos canales de TV abierta evitando la interferencia de estos servicios con las estaciones de televisión. Ambos operan en las mismas bandas.

La idea es utilizar las frecuencias que se han asignado para la transmisión del Sistema Nacional de Televisión Abierta para

combinarlo con un sistema basado en estas normas, pero utilizado para comunicaciones rurales.

5. IEEE 802.22 COMO UNA SOLUCIÓN PARA ÁREAS RURALES.

5.1 Introducción.

En la República Argentina, así como en varios países de América del Sur, los Servicios de Radiodifusión Digital Terrestre han adoptado el estándar japonés con algunas variantes (tales como la sustitución del sistema de compresión MPEG-2 por el MPEG-4).

Este estándar fue desarrollado por la Asociación de Industrias y Empresas de Radiocomunicaciones conocida como ARIB⁴, que promueve el uso eficiente del espectro de frecuencias. El mismo se conoce como Integrated Services Digital Broadcasting⁵ - ISDB / Terrestrial Brazilian Version⁶ ISDB-TB.

Este estándar está compuesto por cuatro normas diferentes diseñadas en función del medio que será utilizado a saber: ISDB-S (sistemas satelitales); ISDB-T (terrestre), ISDB-C (cable) y la radiodifusión móvil 2.6 GHz. Todas ellas radican en la multiplexación de las señales, con una estructura de flujo que permiten transportar señales de Televisión de Alta Definición (HDTV) o de Televisión de Definición Estándar.

El nombre de la norma fue elegida por su similitud con la norma de conocida como RDSI⁷ que es ampliamente conocida por su uso en el transporte de señales de telefonía. Ambas permiten la transmisión si-

4 <http://www.arib.or.jp/english/> - Association of Radio Industries and Businesses.

5 Servicios Digitales Integrados de Radiodifusión.

6 Versión Terrestre Brasileña.

7 Red Digital de Servicios Integrados.

multánea de múltiples canales de datos a través del método de multiplexación.

En el caso de Argentina existe un plan creado en julio de 2013 para la instalación de un Sistema de Televisión Digital Abierta, que cubre un área importante del país y que está estructurado en tres etapas denominadas I, II y III.

El mismo tiene por objetivo cubrir prácticamente la totalidad de las áreas pobladas, dando servicio a 90% de la población nacional. Las etapas I y II comprenden 78 estaciones, las que ya se encuentran funcionando. La etapa III incluye 11 estaciones más que están en construcción.

El uso del espectro de frecuencias para la radiodifusión de televisión ha variado desde las primeras emisiones en blanco y negro a los sistemas actuales de alta definición. Se utilizan básicamente dos bandas: VHF (54 a 88 y 174 a 216 MHz) y la de UHF (512 a 806 MHz).

La distribución geográfica de este sistema de transmisión de señales de televisión con una cobertura tan extensa lleva a proponer que esta inversión ya efectuada y tan importante se pueda utilizar, además, generando un aprovechamiento intenso del espectro de frecuencias.

Precisamente, la norma IEEE 802.22 trabajando en las frecuencias que usa la televisión digital terrestre podría, sin consumir frecuencias adicionales y aprovechando los White Space que dejan estas transmisiones, aprovechar las características de la norma mediante el uso de la tecnología radio cognitiva que la misma utiliza.

5.2 Uso de la IEEE 802.22 en Zonas Rurales.

En nuestros países, el uso intensivo del espectro y la saturación en varias de sus bandas de frecuencia se debe a que las

comunicaciones inalámbricas han sido la única solución viable. En muchos barrios, en especial en zonas suburbanas, las únicas comunicaciones telefónicas posibles son por medio de la telefonía móvil, ya que los mismos por su estructura no pueden ser cableados.

Por otra parte, en nuestro país la televisión en la banda VHF se debería eliminar durante el corriente año (apagón analógico).

Luego para los servicios de televisión digital se utilizarán, tal como se viene haciendo, la banda de UHF. Actualmente, la Autoridad de Aplicación ha licenciado sólo unos pocos canales en las principales ciudades (22 a 36) pero a medida que la nueva tecnología de TV digital permite varios programas de definición estándar en el mismo ancho de banda de un solo canal de alta definición, se produce un ahorro significativo del espectro por lo que se puede obtener gran cantidad de frecuencias libres (canales 38 a 69), principalmente en las ciudades pequeñas.

En Argentina, como en muchos países con grandes zonas rurales, la mayoría de las ciudades se encuentran, en promedio, dentro de un rango de 40 a 80 kilómetros de distancia y la población rural vive alejada no más de esas distancias de los centros poblados.

El estándar IEEE 802.22 tiene un alcance que está dentro de las distancias mencionadas y por otra parte tiene la capacidad de utilizar las frecuencias asignadas al espectro de la Televisión Digital Abierta en la modalidad de radio cognitiva [10, 11].

Ello genera una oportunidad para que esta norma sea considerada en la reasignación del espectro, actualmente en estudio, por parte de la Comisión Nacional de Comunicaciones.

ñalado en este trabajo es cada vez más escaso y congestionado

7. FUTUROS TRABAJOS.

Se estima que se deberá continuar con el estudio de esta recomendación en sus aspectos técnicos, para determinar fundamentalmente sus limitaciones, si ellas existieran.

Se requerirá un relevamiento del equipamiento que el mercado está ofreciendo sobre esta norma y un análisis de las capacidades del mismo. De esa manera se podrá tener una idea de los costos requeridos para cubrir distintas zonas del territorio nacional.

De la misma manera que se hizo con el Proyecto Corral de Lorca se deberá efectuar una prueba de campo para verificar el verdadero rendimiento del equipamiento y la dificultad que requerirá su despliegue.

Las instalaciones de las estaciones base del sistema de televisión digital terrestre, instaladas sobre *shelters*, sin duda pueden ser útiles para el despliegue de parte de los equipos requeridos por la 802.22.

8. FORMACIÓN DE RECURSOS HUMANOS.

Desde el año 2012 en este grupo trabajan alumnos de las carreras de grado de Ingeniería en Informática de las distintas Universidades que participan del proyecto.

Durante el año 2015 se han sumado al proyecto un alumno avanzado de la carrera de Ingeniería en Informática. El mismo ha recibido la beca Estímulo a las Vocaciones Científicas, perteneciente al Programa Estratégico de Investigación y Desarrollo, Plan de Fortalecimiento (Componente de Formación de Recursos Humanos) del Consejo Interuniversitario Nacional. Cabría la posibilidad incluso

que pueda realizar su Trabajo Final de Carrera en algún tema de los que aborda la presente línea de investigación.

9. REFERENCIAS.

[1] Proyecto FONCyT - ANPCyT. PICTO 11-18621. Redes Privadas Comunitarias. Proyecto finalizado y aprobado. Antonio Castro Lechtaler (Director).

[2] García Guibout, J., García Garino C., Castro Lechtaler, A. R., Fusario R., y Sevilla, G., (2007) Physical and Link Layer in Power Line Communications Technologies. *Proceedings of 13th of Argentine Congress on Computer Science*. ISBN 978 - 950 - 656 - 109 - 3. Pág. 56 a 67.

[3] García Guibout, G., García Garino, C., Castro Lechtaler, A. R., Fusario R. y Sevilla, G. (2007) Power Line Communications in the Electric Network. *Proceedings of 13th of Argentine Congress on Computer Science*. ISBN 978 - 950 - 656 - 109 - 3. Pág. 68 a 79.

[4] García Guibout, J., García Garino. C., Castro Lechtaler, A. y Fusario, R., (2008). Transmission voice over 802.11. *Proceedings of 14th of Argentine Congress on Computer Science*. ISBN 978 - 987 - 24611 - 0 - 2. Pág. 307 a 318.

[5] Castro Lechtaler, A., Foti, A., Fusario, R., García Garino, C., y García Guibout, J., (2009) Communication Access to Small and Remote Communities: The Corral de Lorca Project. *Proceedings of 15th of Argentine Congress on Computer Science*. ISBN 978 - 897 - 24068 - 4 - 1. Pag. 1.117 a 1.126.

[6] A. Castro Lechtaler, A. Foti, C. García Garino, J. García Guibout, R. Fusario and A. Arroyo Arzubi. Proyecto Corral de Lorca: Una solución de conectividad a grupos poblacionales pequeños, aislados y distantes de centros urbanos. *Proceedings de la Novena Conferencia Iberoame-*

ricana en Sistemas, Cibernética e Informática: CISCI 2010 - Volume III - ISBN - 13: 978 - 1 - 934272 - 96 - 1. pp. 121 a 127. Orlando, USA. June 2010.

[7] IEEE 802.22 - Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) *Specifications Policies and Procedures for Operation in the TV Bands*.

[8] Gómez, C., (2013). Spectrum Regulation and Policy Officer Radiocommunication Bureau. ITU. Apia, Samoa. www.itu.int/ITU-D/asp/CMS/Events/.../ITU-APT-S3_Cristian_Gomez.pdf

[9] CEPT Report 24. A preliminary assessment of the feasibility of fitting new/future applications/services into non-harmonized spectrum of the digital dividend (namely the so-called "white spaces" between allotments. Report C from CEPT to the European Commission in response to the Mandate on: Technical considerations regarding harmonization options for the Digital Dividend. 1 July 2008.

[10] Mitola, J. and Maguire, G., (1999) Cognitive radio: making software radios more personal. *IEEE Personal Communications Magazine*, Volume 6 Issue 4. Pág. 13 a 18.

[11] Mitola, J. (2000) Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio. Dissertation submitted in partial fulfillment of the degree of Doctor of Technology. Royal Institute of Technology (KTH) - Sweden. Teleinformatics. ISSN 1403 - 5286.

[12] Cordeiro, C., Kiran Challapali, and Dagnachew Birru, Sai Shankar (2006) N. IEEE 802.22: An Introduction to the First Wireless Standard based on Cognitive Radios *Journal of Communications*, Vol. 1, N° 1.

“Herramientas de Software de Simulación para Redes de Comunicaciones”

Juan Antonio Torres¹, Daniel Arias Figueroa², Javier Díaz³

Departamento de Informática – Sede Regional Orán ¹
Facultad de Ciencias Exactas de la Universidad Nacional de Salta
C.I.D.I.A. (Centro de Investigación y Desarrollo de Informática Aplicada)²
Facultad de Ciencias Exactas de la Universidad Nacional de Salta
L.I.N.T.I. (Laboratorio en Nuevas Tecnologías Informáticas)³
Facultad de Informática de la Universidad Nacional de La Plata

juantorresunsa@gmail.com, daaf@cidia.unsa.edu.ar

Resumen

Actualmente los sistemas de redes de comunicación son parte esencial del funcionamiento de las empresas, y una de las tareas más importantes de los diseñadores de redes. El estudio del desempeño de las redes de comunicaciones a través de la simulación es una forma de abordar estas redes de comunicaciones. Las herramientas de simulación Open Source con licencia GNU, son apropiadas para los estudiantes universitarios. En la Sede Regional Orán de la Universidad Nacional de Salta se han ido incorporando estas herramientas, no solo en el ámbito de la red en producción, sino también en el ámbito académico.

Este trabajo presenta parte de los resultados finales de la investigación realizada en el marco de la Especialidad en Seguridad y Redes en la Universidad Nacional de La Plata.

Palabras clave: Open Source, Entorno de simulación, Topologías de red, Laboratorios de red.

Contexto

La línea de investigación se encuentra apoyada por el Departamento de Informática

de la Sede Regional Orán de la Universidad Nacional de Salta, el C.I.D.I.A. (Centro de Investigación y Desarrollo de Informática Aplicada) que depende de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta y el L.I.N.T.I. (Laboratorio en Nuevas Tecnologías Informáticas) – Facultad de Informática de la Universidad Nacional de La Plata.

Introducción

El vertiginoso avance en el campo de la informática, específicamente en las redes de comunicaciones en cuanto a los diferentes dispositivos que se interconectan (celulares, palms, netbook, notebook, tables, iphone, etc.), tráfico y normas de comunicación, ha crecido de manera colosal este último tiempo.

Actualmente los sistemas de computación y redes de comunicación son parte esencial del funcionamiento de las empresas, y una de las tareas más importantes, de los diseñadores de redes. El estudio del desempeño de las redes de comunicaciones a través de la simulación es una forma de abordar estas redes de comunicaciones.

La teoría de simulación ha sido aplicada en una gran variedad de áreas del conocimiento. En los ámbitos educativos se puede decir que la mayoría de las actividades de aprendizaje casi siempre están en entidades de

simulación, en todo momento profesor y alumno están trabajando con hipótesis y supuestos ya que en pocas ocasiones el profesor sale del aula y se va con sus alumnos al mundo exterior para explicar y demostrar teoremas, leyes, hipótesis, etc. Muchas propuestas de trabajos están enunciadas en base a una semirealidad.

Por lo antes expuesto y, considerando que existen en el mercado distintas herramientas para la simulación de redes de comunicaciones, este trabajo intenta clarificar las distintas líneas de acción de estos simuladores con licencia GNU GPL.

Líneas de Investigación, Desarrollo e Innovación

El objetivo de este trabajo es describir las características más importantes y requisitos necesarios para el funcionamiento de los simuladores de redes de comunicaciones utilizados para el aprendizaje, modelado y análisis, con licencias GNU GPL.

Tipos de Simuladores

Existen distintos tipos de simulaciones que buscan imitar el comportamiento del mundo real, y no utilizan un dispositivo informático como por ejemplo: las actividades de dramatización, juegos de rol, maniquí de instrucción para prácticas de enfermería, etc. Por ello, es conveniente revisar algunas definiciones de simulación que las avalen.

*“La simulación es una técnica que permite reproducir la esencia de un fenómeno sin reproducirse el fenómeno en si”*¹

“La simulación es un método de enseñanza que propone acercar a los alumnos a situaciones y elementos similares a la realidad, pero en forma artificial, a fin de entrenarlos en habilidades prácticas y

¹ KLETON, David W. Simulation Modeling and Analysis. Tercera Edición. Mc Graw Hill. 2000

*operativas cuando las encaran en el mundo real”*².

Este trabajo evalúa los SIMULADORES DE REDES DE COMUNICACIONES, por lo tanto se revisarán las definiciones que lo avalen:

*“La simulación de redes telemáticas es la metodología que implementa una red en una computadora para su posterior evaluación. Permite probar escenarios que son difíciles o costosos de simular en entornos reales”*³

*“La simulación de redes es sin duda la metodología de evaluación más predominante en el área de las redes telemáticas. Una red telemática estará compuesta por la descripción del comportamiento de los nodos que la componen y de sus canales de comunicación, así como de un conjunto de parámetros que permitan configurar de forma dinámica el comportamiento de los nodos y los canales de comunicación”*⁴

La definición propuesta para este trabajo es la enunciada a continuación:

“Las herramientas de software de simulación para redes de comunicaciones intentan imitar el comportamiento real de equipos físicos interconectados”

² Davini, Cristina; **Métodos de Enseñanza** ; Ed. Santillana; 2010

³ S. Siraj, A. K. Gupta, and R. Badgujar, **“Network Simulation Tools Survey”**, International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE 2012), vol. 1, pp. 201–210, junio 2012

⁴ E. Weingartner, H. vom Lehn, and K. Wehrle, **“A Performance Comparison of Recent Network Simulators,”** in IEEE International Conference on Communications (ICC’09), pp. 1–5, junio 2009

Simuladores de Redes de Comunicaciones

Una de las tareas más importantes de los diseñadores de redes de comunicaciones es el estudio del desempeño de las redes de computadoras. Esto es debido a que una mala decisión de diseño puede afectar fuertemente el desempeño de la red y provocar pérdidas económicas para la empresa.

Por otro lado, los simuladores de redes de comunicaciones en la enseñanza son elementos activos de aprendizaje, ya que estas herramientas facilitan el estudio y comprensión de conceptos y fundamentos teóricos.

La realización de prácticas de redes de datos sobre equipos reales, tropieza con la dificultad que los laboratorios, en general, están acotados en la cantidad y variedad de equipos, por lo que hay pocas posibilidades de variar el diseño y muchas veces se debe trabajar con una sola topología existente. Además de encontrarse las dificultades propias de los laboratorios, como son la coordinación con otras personas que hacen uso del mismo, el mantenimiento y la cantidad de personas que puede contener.

Si bien un simulador no puede sustituir el trabajo directo con equipos, puede proveer en cambio: facilidad de acceso, manejo de diversas topologías, equipos y protocolos, rapidez en el armado, trabajo con diferentes tipos de escenarios, algunos de estos escenarios pueden ser configurados erróneamente o incompletos para corregirlos y, un punto clave, la visualización gráfica. Aunque no son reales, imitan de cerca la realidad. Otro punto a favor de los simuladores de redes de comunicaciones son las prácticas domiciliarias, en cualquier horario y al ritmo propio del estudiante.

Ventajas

- ✓ Aprendizaje por descubrimiento.
- ✓ Promoción de la creatividad.
- ✓ Ahorro de tiempo y dinero.
- ✓ Confianza.

- ✓ Enseñanza individualizada y domiciliaria.
- ✓ Autoevaluación.
- ✓ Evita riesgos (no se expone a situaciones de peligro).
- ✓ Repetición.
- ✓ Escenarios complejos.

Resultados alcanzados

Si bien existen una variedad de simuladores en cuanto a la cantidad de dispositivos virtuales y protocolos que pueden soportar. La mayoría buscan la formación del usuario y/o investigación en materia de redes informática. Los simuladores tienden a servir de apoyo en las prácticas de laboratorio para formación académica universitaria.

Algunos de los simuladores son básicos y no tienen versatilidad para cambiar la topología existente como **JIMSIM** y **TOGGIT**. Otros no tan básicos, trabajan con topologías definida por el usuario y en general con el protocolo TCP/IP como es el caso de **KivaNS**, **Marionnet**, **NET-Simulator** y **PSimulator2**, todos ellos están orientado a la formación académica y son patrocinados por las universidades (a excepción de Net Simulator), nacidos como trabajo final de tesis o investigación. También se encuentran aquellos simuladores que son productos de grandes corporaciones o instituciones, ofrecen gran abanico de protocolos, laboratorios, apoyo, amplia documentación, etc. como los son **CORE**, **GNS3** **OMNET++**, **OPNET**. Estos simuladores al poseer licencia GNU GPL, facilitan la investigación y colaboración en dichos proyectos. Si bien, **OPNET** posee Licencia propietaria brinda licencia gratuita renovable cada seis meses para entornos educativos, lo cual lo hace interesante.

Un simulador destacable es **GNS3**, por el tipo de licencia (GNU GPL v2), la cantidad de idiomas (19), su cercanía a los dispositivos

reales y la capacidad de integración con otros tipos de software como MVware (o VirtualBox), Dynamips, Qemu, putty y Wireshark.

La dificultad que emergió en el desarrollo de este trabajo ha sido obtener bibliografía, en tanto han sido muy valiosos los distintos papers encontrados en la web.

A futuro este trabajo servirá como material de recopilación de las distintas herramientas de simulación de redes de comunicaciones con Licencia GNU GPL.

Formación de Recursos Humanos

La formación de postgrado producto de la investigación se detalla a continuación.

- Especialidad en Seguridad y Redes – UNLP. “Virtualización para la enseñanza de redes IP”. Expte. 3300-3489/11. Miguel A. Aguirre.
- Especialidad en Seguridad y Redes – UNLP. “Herramientas de Software de Simulación para Redes de Comunicaciones”. Expte. 3300-5305/12. Juan Antonio Torres. (Aprobada en julio de 2015).

Referencias

KLETON, David W. **Simulation Modeling and Analysis**. Tercera Edición. Mc Graw Hill. 2000

Davini, Cristina; **Métodos de Enseñanza**; Ed. Santillana; 2010

S. Siraj, A. K. Gupta, and R. Badgujar, “**Network Simulation Tools Survey**”, International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE 2012), vol. 1, pp. 201–210, junio 2012

Edith Chacón Osorio, Wilder Eduardo Castellanos Hernández Mónica, “**UTILIZACIÓN DE HERRAMIENTAS SOFTWARE PARA EL MODELADO Y LA SIMULACIÓN DE REDES DE COMUNICACIONES**”.

P. Belzarena, v. Gonzalez- Barbone, “**INCORPORACION DE UN SIMULADOR GRÁFICO DE REDES EN UN OBJETO DE APRENDIZAJE REUTILIZABLE**”,

Eduardo Zornoza Martínez, “**APRENDIZAJE CON SIMULADORES. APLICACIÓN A LAS REDES DE COMUNICACIONES**”,

Francisco Javier Ruiz, David Fernández, Ana B. García, Fernando Muñoz, Luis Bellido, José I. Moreno, “**IMPLANTACIÓN DE UN LABORATORIO DOCENTE PARA REDES DE COMUNICACIONES**”.

José Manuel Ruiz Gutiérrez, “**LA SIMULACIÓN COMO INSTRUMENTO DE APRENDIZAJE**” (Evaluación de Herramientas y estrategias de aplicación en el aula)

SusanChenMok, **SIMULACION DE REDES DE COMPUTADORAS**.

Evaluación de performance en Redes Definidas por Software para Centro de Datos – Hipótesis de Trabajo

Diego Bolatti, Ricardo Calcagno,
Carlos Cuevas, Sergio Gramajo, Reinaldo Scappini

Grupo Ingeniería en Sistemas de Información, Universidad Tecnológica
Nacional Facultad Regional Resistencia,
French 414, (3500) Chaco, Argentina. Tel. 362-4432683
{dbolatti, rcalcagno, cac, sergio, rscappini}@frre.utn.edu.ar

Resumen

El análisis de infraestructuras de redes y centros de datos orientados a servicios se vuelve un elemento crítico en las empresas de telecomunicaciones. A medida que aumentan los servicios y el software las exigencias de usuarios para contar con disponibilidad, garantía, calidad y seguridad de la información son mayores.

Este contexto hace que la evaluación de los DATACENTERS respecto a numerosos criterios como servidores, virtualización, costos, uso de recursos, la variabilidad de las necesidades de los usuarios, y las posibilidades de crecimiento, entre otros, sean tenidas en cuenta.

En este trabajo se evaluarán estos aspectos sobre nuevas tecnologías como son redes de próxima generación y Redes Definidas por Software (SDN) aplicadas en DATACENTERS.

Para finalizar, se desarrollará un modelo de evaluación de performance que ayude a los administradores de red a tomar decisiones en base a atributos críticos (tipos de tráfico, servicios, usuarios, etc.) identificados previamente.

Palabras clave: Redes Definidas por Software, Análisis de Tráfico, Data Centers.

Contexto

Este proyecto está inserto en una línea de I/D presentada en la Universidad Tecnológica Nacional con código: UTN-2422. Título: “*Modelo para la evaluación de performance mediante identificación de tráfico y atributos críticos en Redes Definidas por Software*”. Dicho proyecto se lleva a cabo en el ámbito del Dpto. de Ingeniería en Sistemas de Información perteneciente a la Facultad Regional Resistencia de la Universidad Tecnológica Nacional.

Introducción

Un **DATACENTER** es un centro de procesamiento de datos, una instalación empleada para albergar un sistema de información de componentes asociados, como telecomunicaciones y los sistemas de almacenamientos. Generalmente incluyen diferentes dispositivos de seguridad para permitir que los equipos tengan el mejor nivel de rendimiento con la máxima disponibilidad del sistema.

Un DATACENTER ofrece varios niveles de resguardo, en forma de fuentes de energía de backup y conexiones adicionales de comunicación, que puede no ser utilizada hasta que pase algún problema en el sistema primario donde el principal objetivo de estas

instalaciones es ejecutar las aplicaciones centrales del negocio y almacenar datos operativos, donde ofrece las aplicaciones más tradicionales que es el sistema de software corporativo como Enterprise Resource Planning (ERP) y Customer Relationship Management (CRM), entre otros.

Los componentes más comunes son conmutadores, firewalls, gateways VPN, routers y servidores de datos, de archivos, de aplicaciones (gestión, web, etc.) y middleware, todo en hardware físico o en plataformas consolidadas y virtualizadas.

Actualmente se implementan redes de datos de alta velocidad y redundantes que interconectan los equipos que componen el DATACENTER. Esta interconexión tiene por objetivo mejorar los tiempos de respuesta de los servidores y de las redes de almacenamiento (SAN – Storage Área Network) y desarrollar sistemas de almacenamiento de datos con mayores capacidades y velocidades de respuesta que dan soporte al servicio prestado por el mismo.

Las **Redes Definidas por Software** [9] [10] proponen un modelo para cubrir nuevas demandas de usuarios y organizaciones. Ésta es una arquitectura de red emergente, donde el control de la infraestructura de red está desacoplado del reenvío de datos y, a su vez, es directamente programable.

La inteligencia de red es (lógicamente) centralizada en controladores SDN basados en software que mantienen una visión global de la red. Como resultado, las organizaciones controlan la red independiente del proveedor en un único punto lógico lo que simplifica, en gran medida, el diseño de la red y su operación. A su vez, se simplifica la gestión de los dispositivos de red debido a que ahora no se tienen que entender y procesar miles de normas de protocolo, sino simplemente aceptar instrucciones de los controladores centralizados de SDN (ver Fig. 1).

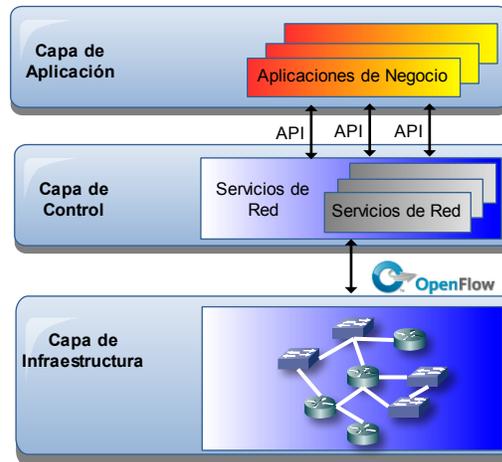


Figura 1. Arquitectura de Software-Defined Network

Tal vez lo más importante de esta nueva visión de red, es que se pueda configurar mediante programación esta abstracción simplificada de la red en lugar de tener que configurar manualmente múltiples dispositivos. Además, aprovechando la inteligencia centralizada del controlador SDN, se puede alterar el comportamiento de la red en tiempo real y desarrollar nuevas aplicaciones y servicios de red ágilmente, lo que mejora sustancialmente las posibilidades [11][12].

Para poder concretar la nueva arquitectura de redes fue necesario crear y estandarizar una interfaz de comunicaciones entre el control y el reenvío de datos. Para ello se creó el protocolo OpenFlow [13] que permite el acceso directo a la gestión de datos de reenvío en dispositivos de red como switches y routers, tanto físicos como virtuales y de un modo abierto. Esto contrasta con las arquitecturas de redes tradicionales donde los dispositivos de red son monolíticos y cerrados.

Ningún otro protocolo estándar tiene la funcionalidad y finalidad de OpenFlow que transfiere el control de los dispositivos de red a la lógica del software de control (ver Fig. 2).

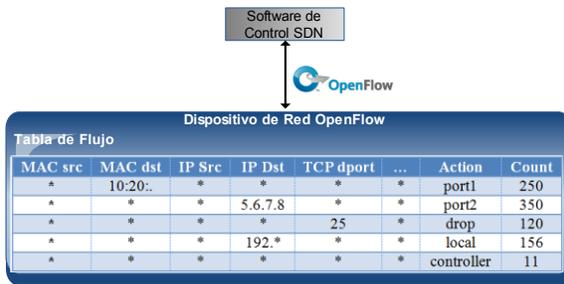


Figura 2. Instrucciones OpenFlow

OpenFlow utiliza el concepto de flujos para identificar el tráfico de red basado en reglas estáticas o dinámicas programadas por el software de control de SDN. También permite definir el modo en que el tráfico debe rutearse a través de los dispositivos. Debido a que OpenFlow permite ser programado sobre una base de flujo, una arquitectura basada SDN-OpenFlow proporciona un control muy granular de la infraestructura de red, lo que permite responder a los cambios en tiempo real en la aplicación, el usuario e inclusive en los niveles de sesión.

Una característica particular de SDN basada en OpenFlow es que se puede implementar en las redes existentes, tanto físicas como virtuales. Los dispositivos de red pueden realizar el reenvío basado en OpenFlow, así como el reenvío tradicional, lo que hace que sea muy fácil para las organizaciones introducir progresivamente esta tecnología, incluso en los entornos de red de múltiples proveedores.

Líneas de investigación y desarrollo

En el proyecto “Modelo para la evaluación de performance mediante identificación de tráfico y atributos críticos en Redes Definidas por Software” se propone el análisis de la arquitectura y estándar de SDN. Además del diseño e implementación de un sistema de soporte a las decisiones con Información Lingüística [14][15][16] para evaluar datos cualitativos de SDN y su uso. Este sistema permite que múltiples expertos puedan participar conjuntamente dando sus

puntos de vista que ayuden a la toma de decisiones.

Los aspectos cuantitativos del modelo se realizan mediante simuladores y creaciones de escenarios de comparación con las redes tradicionales.

Este aspecto particular del proceso de investigación está destinado a los DATA-CENTER y su uso para SDN

La relación entre ambos tópicos es la virtualización como herramienta de administración de los DATACENTERS, y que mediante la implementación de SDN se lograría una mejor administración de los recursos de comunicaciones dentro de mismo. Esta mejora mencionada está definida por la configuración del flujo de datos que se puede lograr de los equipos de conmutación virtuales internos que comunican las máquinas virtuales dentro de DATACENTER.

Resultados y objetivos

Este proyecto es reciente pero se han obtenido algunos resultados teóricos en el análisis y comparación de las arquitecturas de SDN con las redes tradicionales y los modelos de simulación de tráfico. Basado en estas premisas y partiendo de la arquitectura general propuesta en [17]. Ver Figura 3

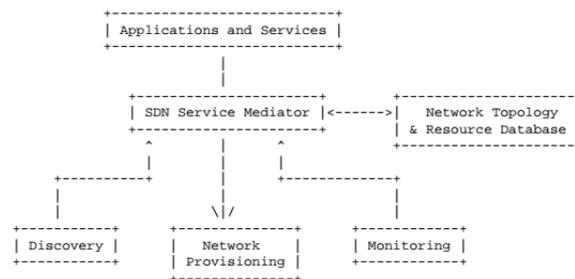


Figure 3: Proposed SDN Architecture

Figura 3. Arquitectura de Software-Defined Network propuesta para DATACENTER [17]

En el bloque funcional **Network Topology & Resource Database** el proceso

Network Provisioning, destinado a suministrar los recursos necesarios en función de los requerimientos de los usuarios y/o aplicaciones, podemos identificar dos hipótesis de trabajo concretas:

1. Actividades relativas al modelado y simulación de la red propiamente dicha
2. Identificación atributos críticos como insumo para el modelo de toma de decisiones

En particular, el trabajo de definición de atributos se puede enfocar en el módulo de Aprovisionamiento de Red, identificando los atributos que es necesario proveer al **Service Mediator** para crear o modificar los perfiles de tráfico y, consecuentemente, las conexiones de ingeniería de tráfico. Se debe considerar, por ejemplo, si es conveniente encontrar atributos comunes a varias tecnologías, versus seleccionar varios conjuntos de atributos específicos para cada tecnología en particular (ej. VLAN, MPLS, wireless). Asimismo, otro aspecto a considerar es la relación del módulo de Aprovisionamiento con los estándares y herramientas de SDN mencionadas en el trabajo (OpenFlow, Net-Conf, etc)

En el curso del trabajo de investigación, se puede evaluar la conveniencia de incorporar otros módulos con sus respectivos atributos para el modelado de toma de decisiones, como ser Monitoreo y Descubrimiento. De más está decir que, a priori y en este estadio tan temprano del trabajo, de todas maneras surge como un aspecto interesante la vinculación de los tres módulos al proceso de toma de decisiones.

Formación de Recursos Humanos

La formación de recursos humanos es la siguiente:

Formación de becarios:

Este año se incorporarán dos becarios alumnos avanzados de la carrera de ingeniería en sistemas de información y un becario graduado de iniciación a la investigación. Esto hará posible fomentar la actividad de investigación en alumnos que están próximos a recibirse y graduados jóvenes estimulando la actividad de investigación.

Formación de postgrado:

A partir de las líneas de investigación desarrolladas en el proyecto se prevé que el Ing. Carlos Cuevas finalice su Maestría en Redes de la Universidad de La Plata mediante una tesis vinculada a este proyecto.

Así mismo servirá de base para la investigación y formulación de la tesis de Maestría en Administración de Negocios para el Ingeniero Diego Bolatti. Carrera de postgrado cursada en UTN Facultad Regional Resistencia.

Equipo de trabajo:

La estructura del equipo de trabajo es la siguiente:

Director	<ul style="list-style-type: none"> • Dr. Ing. Sergio Gra-majo
Investigadores: (En orden alfabético)	<ul style="list-style-type: none"> • Ing. Diego Bolatti • Ing. Ricardo Calcagno • Ing. Carlos Cuevas • Ing. Reinaldo José Ramón Scappini
Colaborador Externo	<ul style="list-style-type: none"> • Dr. Luis Martínez López (Catedrático Universidad de Jaén, España) • Dra. Macarena Espinilla Estévez (Investigadora Universidad de Jaén)

Referencias

1. Vegesna, S.: IP Quality of Service. Cisco Press, Indianapolis (2001)
2. Ash, G.: Traffic engineering and QoS optimization technology overview. Traffic Engineering and QoS Optimization of Integrated Voice & Data Networks, pp. 409-428. Morgan Kaufmann, San Francisco (2007)
3. Leland, W., Taqqu, M., Willinger, W., Wilson, D.: On the self-similar nature of ethernet traffic (extended version). IEEE/ACM Transactions on Networking 2, (1994)
4. Kihong Park, W.W.: Self-Similar Network Traffic and Performance Evaluation. John Wiley & Sons, Inc., New York (2000)
5. Osborne, E., Simha, A.: Traffic Engineering with MPLS (2002)
6. Szigeti, T., Hattingh, C.: End-to-End QoS Network Design: Quality of Service in LANs, WANs, and VPNs. Cisco Press, Indianapolis (2004)
7. Evans, J., Filsfils, C.: Deploying IP and MPLS QOS for Multiservice Networks. Theory and Practice. Elsevier (2007)
8. Jeong, S.H., Owen, H., Copeland, J., Sokol, J.: QoS support for UDP/TCP based networks. Computer Communications 24, 64-77 (2001)
9. Kobayashi, M., Seetharaman, S., Parulkar, G., Appenzeller, G., Little, J., van Reijendam, J., Weissmann, P., McKeown, N.: Maturing of OpenFlow and Software-defined Networking through deployments. Computer Networks 61, 151-175 (2014)
10. Open-Networking-Foundation: Software-Defined Networking: The New Norm for Networks. <https://www.opennetworking.org/sdn-resources/sdn-library/whitepapers>, (2013)
11. Das, S., Talayco, D., Sherwood, R.: Chapter 17 - Software-Defined Networking and OpenFlow. In: DeCusatis, C. (ed.) Handbook of Fiber Optic Data Communication (Fourth Edition), pp. 427-445. Academic Press, Oxford (2013)
12. Paul, S., Jain, R., Samaka, M., Pan, J.: Application delivery in multi-cloud environments using software defined networking. Computer Networks (2014)
13. Open-Networking-Foundation: Open-Flow Switch Specification. <https://www.opennetworking.org/sdn-resources/onf-specifications/openflow>, (2013)
14. Espinilla, M., Liu, J., Martínez, L.: An extended hierarchical linguistic model for decision-making problems. Computational Intelligence 27, 489-512 (2011)
15. Herrera, F., Herrera-Viedma, E.: Linguistic decision analysis: Steps for solving decision problems under linguistic information. Fuzzy Sets and Systems 115, 67-82 (2000)
16. Zadeh, L.: The concept of a linguistic variable and its application to approximate reasoning. Part I, Part II and Part III. Information Sciences 199-249, 301-357, 143-180 (1975)
17. P. Pan-T. Nadeau: Software-Defined Network (SDN) Problem Statement and Use Cases for Data Center Applications. Network Working Group Internet-Draft, 7-8 (Sep 2012).

Controlador DNP3 para la CIAA

Sebastián Tobar, Joel Noguera, Carlos Taffernaberry, Gustavo Mercado.

GridTICs – Grupo en Tecnologías de la Información y las Comunicaciones /
Departamento de Electrónica / Facultad Regional Mendoza / Universidad
Tecnológica Nacional

Rodríguez 273 Capital – Mendoza, +54 261 5244563

{sebastian.tobar, joel.noguera, carlos.taffernaberry, gustavo.mercado,}@gridtics.frm.utn.edu.ar

Resumen

La CIAA [1] (Computadora Industrial Abierta Argentina) tiene las características necesarias para utilizarse como RTU (Remote Terminal Unit - Unidad Terminal Remota) en sistemas SCADA [2] (Supervisory Control And Data Acquisition – Supervisión Control y Adquisición de Datos). Sin embargo, una gran limitación para esta aplicación radica en que actualmente soporta un único protocolo de comunicación: Modbus. Si bien este protocolo es, probablemente, el más utilizado históricamente en la industria, es un protocolo antiguo que carece de numerosas características altamente requeridas en ambientes industriales modernos, tales como:

- Eficiencia en el uso del ancho de banda
- Soporte de estampas de tiempo
- Seguridad
- Reportes por excepción

Todas estas características son propias del protocolo DNP3 [3], el cual se utiliza ampliamente en la industria argentina, principalmente en la industria eléctrica (generación, transporte y distribución) y con requerimientos crecientes en la industria petrolera y agua y saneamiento.

El presente proyecto desarrollará un controlador DNP3 nativo para la CIAA, siguiendo la filosofía de código abierto de la plataforma. Esta implementación, permitirá la integración de la CIAA a sistemas eléctricos inteligentes [4] (smartgrid) como así también sistemas de agua y saneamiento inteligentes (smartwater).

Palabras clave: dnp, ciaa, rtu, scada, Smart grid, Smart city, Internet de las Cosas

Contexto

El Ministerio de Ciencia, Tecnología e Innovación Productiva, a través de la Subsecretaría de Políticas en Ciencia, Tecnología e Innovación Productiva y el Programa Nacional I+DEL, convocó a la presentación de propuestas destinadas a la adjudicación de “Proyectos de Innovación a partir de la adopción de la Computadora Industrial Abierta Argentina en productos y procesos industriales”. El presente trabajo fue presentado y seleccionado en esta convocatoria según resolución MinCyT 613/15.

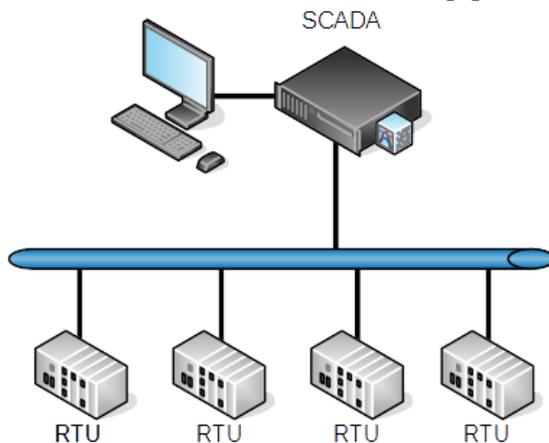
Dentro del GridTICs, este proyecto es un paso más dentro de la investigación de

la incursión de las TICs en las redes eléctricas [5]. Principalmente articula con el proyecto "Red Inteligente Ciudad General San Martín, Mendoza" [6], aprobado y financiado por la convocatoria FITS 2013 UREE - Uso Racional y Eficiente de la Energía de la Agencia Nacional de Promoción Científica y Tecnológica.

Introducción

SCADA es una tecnología que permite a un usuario recopilar datos y enviar instrucciones de control hacia una o más instalaciones distantes. De manera más específica, un sistema SCADA permite a un operador ubicado en una sala de control central, enviar consignas a un proceso, abrir o cerrar válvulas, abrir o cerrar interruptores, monitorear alarmas y obtener mediciones de procesos distribuidos tales como campos petroleros, oleoductos, sistemas de agua potable, sistemas de gas natural, centrales hidroeléctricas, líneas de transmisión, etc.

El sistema SCADA abarca la recolección de información a través de RTU, la transferencia de esta información a un sitio central, el análisis y presentación gráfica de esta información y finalmente en envío de acciones de control nuevamente hacia la RTU [7].



Para el funcionamiento de estos sistemas, es indispensable el uso de sistemas de comunicaciones confiables, que garanticen la correcta supervisión y control de los procesos.

El protocolo DNP3, desarrollado originalmente por Harris, Distributed Automation Products, fue diseñado específicamente para aplicaciones SCADA. Es un protocolo de propósitos generales, basado en las reglas de diseño plasmadas en las primeras partes del estándar IEC 60870, específicamente 60870-5-1 [8], 60870-5-2 [9] y 60870-5-3 [10]. La serie de normas IEC 60870-5 son reconocidas como estándares internacionales para los protocolos de comunicación de sistemas SCADA en la industria eléctrica.

Desde su creación para la industria de la distribución eléctrica en Norteamérica, DNP3 ha ampliado su aceptación industrial y geográfica. El protocolo es soportado por un gran número de fabricantes y usuarios en infraestructura eléctrica, agua, gas, petróleo, seguridad y otras industrias en Norteamérica, Sudamérica, Sudáfrica, Asia, Australia y Nueva Zelanda.

Características fundamentales:

- Protocolo abierto, disponible para ser usado por cualquier fabricante o usuario.
- Diseñado para comunicación eficiente de control y datos.
- Ampliamente soportado por fabricantes.
- Documentación completa y extensa.

Líneas de Investigación, Desarrollo e Innovación

A partir de un estudio detallado del estándar IEEE 1815-2012 - IEEE Standard for Electric Power Systems

Communications-Distributed Network Protocol (DNP3) se implementará un controlador de este protocolo para la CIAA. El mismo está definido para su utilización sobre interfaces seriales o conexiones TCP/IP, y cada dispositivo se puede comportar como maestro o esclavo. En el presente proyecto se implementará un controlador esclavo con capacidad de ser utilizado en el puerto serial de la CIAA. La funcionalidad detallada de este controlador se definirá en la etapa de Análisis de protocolo, en función del estudio realizado del estándar IEEE 1815.

Con respecto al entorno de desarrollo a utilizar para la implementación, en concordancia con la filosofía de la CIAA, se prevé la utilización de software libre en todos los casos en que sea posible. Este software puede incluir (pero no limitarse) a:

- gcc
- ld
- gdb
- make
- Eclipse

Es conveniente mencionar que el desarrollo objeto de este proyecto, será liberado a la comunidad como código abierto y libre.

Para el desarrollo del controlador propiamente dicho, se consideraron varios modelos de desarrollo. Entre ellos podemos mencionar el modelo cascada, modelo espiral, iterativo, ágil, extremo, etc. En función de estas opciones, y debido a la necesidad de obtener un software funcional en el menor tiempo posible, se opta por un modelo de desarrollo ágil, el cual brindará los siguientes beneficios:

- El software se va construyendo de manera incremental
- En cada paso, el software es totalmente funcional
- Se evita desarrollar funcionalidad innecesaria

Los ensayos del protocolo se realizarán utilizando el hardware y software real a ser utilizado en la industria. Para este proyecto, se utilizará el software SCADA GE iFix el cual tiene amplios antecedentes de utilización en la industria eléctrica, alimenticia y agua y saneamiento, entre otras. En conjunto con el SCADA GE iFix, se utilizará el controlador DNP3 suministrado por el fabricante, el cual se utilizará como sistema maestro que interrogará al controlador esclavo implementado en la CIAA.

La utilización de un software probado en la industria garantiza confiabilidad en las pruebas del protocolo. Sin embargo, además de estos ensayos, se deberán llevar adelante las pruebas recomendadas por el DNP3 Users Group [11], las cuales definirán el nivel de conformidad del desarrollo con los estándares del protocolo.

Resultados y Objetivos

El objetivo principal del proyecto es ampliar la cantidad de industrias que pueden utilizar la CIAA.

Como objetivos secundarios podemos mencionar:

- Inserción de la CIAA en las industrias de servicios públicos: Agua, saneamiento, gas, etc.
- Inserción de la CIAA en la industria energética: Petróleo, electricidad.
- Sustitución de equipamiento importado de alto costo.
- Generación de tecnología y know-how local aplicable a nuevas tendencias internacionales: smartgrid (redes eléctricas inteligentes), smartwater (redes de agua inteligentes).

- Integración de la CIAA con productos industriales de estantería.

La CIAA, al igual que cualquier desarrollo en curso, evoluciona de manera permanente. Este proyecto pretende ser parte de esta evolución ampliando el campo de industrias en las cuales la CIAA puede brindar soluciones, mejorando de manera importante sus capacidades de comunicación. Además, vale mencionar que al proponer un desarrollo de código libre, el mismo estará disponible para toda la comunidad de desarrolladores, estudiantes y usuarios, contribuyendo de esta manera el avance y expansión de la CIAA.

Finalmente, con la inminente implementación a nivel mundial de sistemas eléctricos inteligentes (smartgrid) como así también sistemas de agua y saneamiento inteligentes (smartwater), resulta conveniente disponer de una plataforma abierta y de bajo costo en condiciones de integrarse a estos sistemas los cuales demandarán una gran cantidad de recursos tanto humanos como materiales.

Formación de Recursos Humanos

La ejecución de este proyecto consolida la incursión del GridTICs en la industria eléctrica, posibilitando de esta manera la generación de know how local para afrontar el desarrollo y problemáticas del área.

En el proyecto participa un investigador formado, un becario graduado en Ing. Electrónica con amplia experiencia en el mercado eléctrico, un graduado en Ing. en Sistemas y un estudiante de la Especialización de Redes de Datos, dictada por GridTICs.

Referencias

- [1] ACSE / CADIEEL (2013). Disponible: www.proyecto-ciaa.com.ar
- [2] D. J. Gaushell, H. T. Darlington, "Supervisory control and data acquisition", Proc. IEEE, vol. 75, pp. 1645 – 1658, Jun. 2005.
- [3] "IEEE Standard for Electric Power Systems Communications-Distributed Network Protocol (DNP3)", IEEE Standard 1815, 2012.
- [4] "Assessment of Demand Response & Advanced Metering", United States Federal Energy Regulatory Commission, 2008.
- [5] B. Panajotovic , M. Jankovic , B. Odadzic, "ICT and smart grid", in 10th International Conference on Telecommunication in Modern Satellite Cable and Broadcasting Services (TELSIKS), 2011, pp 118-121.
- [6] Gustavo Mercado et al "SG-SM - Smart Grid San Martin - Red de Distribución y Generación de Energía Inteligente en Ciudad Gral San Martin – Mendoza", Anales del WICC 2015, XVII Workshop de Investigadores en Ciencias de la Computación, Universidad Nacional de Salta, 16 y 17 de abril de 2015
- [7] "IEEE Standard for SCADA and Automation Systems", IEEE Standard C37.1, 2007.
- [8] "Transmission Frame Formats", IEC 60870-5-1, 1990.
- [9] "Data Link Transmission Services", IEC 60870-5-2, 1992.
- [10] "General Structure of Application Data", IEC 60870-5-3, 1992.
- [11] DNP Users Group, Disponible: www.dnp.org

RED SIPIA-LP

Estudio de mecanismos de bajo consumo energético aplicados a Red de Sensores Inalámbricos en el ámbito de Agricultura de Precisión.

Ana Diedrichs, Carlos Taffernaberry, Gustavo Mercado, Guillermo Grunwaldt, Matias Pecchia, Germán Tabacchi, Matías Gonzalez, Nicolás Altamiranda.

GridTICs – Grupo en Tecnologías de la Información y las Comunicaciones
Departamento de Electrónica / Facultad Reg. Mendoza / UTN Rodríguez 273, Capital – Mendoza,
{ana.diedrichs, gustavo.mercado, carlos.taffernaberry}@gridtics.frm.utn.edu.ar

Resumen

El proyecto SIPIA-LP tiene como objetivo la optimización del uso energético de una red inalámbrica de sensores (WSN), desarrollada en el contexto de un Proyecto de Investigación y Desarrollo anterior, e instalada en una finca para el monitoreo de parámetros ambientales para el estudio de heladas agronómicas. La WSN está basada en la norma IEEE 802.15.4, que establece el comportamiento a nivel de capa física y de enlace, utilizando la banda libre de 2.4 Ghz. Para la interoperabilidad de estas redes y nodos entre sí, es fundamental la evaluación y estudio de distintos protocolos de comunicación de la Internet de las Cosas, como 6LoWPAN, que permitiría que cada nodo contara con una dirección IPV6. Se está trabajando en el laboratorio en la construcción de una cama de pruebas de sensores inalámbricos en el laboratorio, para comprobar el funcionamiento de los protocolos, métricas de conectividad, y estabilidad del sistema operativo de los nodos (Contiki OS).

Palabras Clave: WSN, 6LoWPAN, 802.15.4, Internet de las Cosas, Agricultura de Precisión, Bajo Consumo

Contexto

El presente trabajo está inserto en el proyecto de investigación acreditado por la Universidad Tecnológica Nacional código UTI3646TC denominado RED SIPIA-LP: estudio de mecanismos de bajo consumo para aplicar a una red de sensores inalámbricos en el ámbito de la agricultura de precisión.

Introducción

La agricultura de precisión consiste en el uso de sistemas de información basados en diversas tecnologías aplicadas al ámbito de la producción agraria y una de las áreas donde puede implantarse con mayor rapidez las tecnologías de sensado

inalámbricas. Por ejemplo, las redes de sensores favorecen una reducción en el consumo de agua y pesticidas, contribuyendo a la preservación del entorno. Adicionalmente, pueden alertar sobre la llegada de heladas, así como ayudar en el trabajo de las cosechadoras. Por medio de sensores estratégicamente situados, se pueden monitorizar parámetros tales como la temperatura o la humedad de las hojas, con el fin de detectar rápidamente situaciones adversas y desencadenar los tratamientos apropiados. Ya existen varios emprendimientos comerciales que venden estas soluciones y algunas incluyen soluciones de sensado inalámbrico, como: Verdtech S.A. [1], Adcon Telemetry [2], Crossbow's technology [3], D2ARS (“Diseño y Desarrollo de Aplicaciones de Redes de Sensores” [4]), Libelium [5].

Redes de Sensores

Un conjunto de motes comunicados entre sí es lo que conocemos como una red de sensores inalámbrica (Wireless Sensor Network o WSN) [6]. Los motes, nodos de la WSN, constan de una unidad de procesamiento con un poder de cómputo mínimo, memoria, una unidad de comunicación inalámbrica, baterías y uno o varios dispositivos de sensado que capturan parámetros como temperatura, aceleración, humedad, etc. Son dispositivos sensores multifuncionales y multipropósito de bajo costo que operan con poca energía, de un tamaño pequeño, y de una capacidad de comunicación inalámbrica a corta distancia. Los nodos sensores pueden variar según el fabricante en capacidades de cómputo, memoria, interfaces de comunicación, etc.

La distribución de los mismos puede ser aleatoria o planeada, lo cual permite su uso en prácticamente en cualquier ambiente físico. Esta característica es provista por un conjunto de protocolos y algoritmos específicos para WSN. Otro de los beneficios que caracteriza a una red de sensores es que sus nodos pueden trabajar de modo

cooperativo, aumentando sus posibilidades de aplicación, por ejemplo, procesando los datos capturados antes de comunicarlos al usuario final o concentrador, o utilizando funciones de auto-configuración para ser más tolerantes a las fallas.

Las redes inalámbricas de sensores proveen la flexibilidad necesaria para disminuir tiempos de instalación, recolección de datos y mantenimiento, si se les compara con recolectores de datos (data loggers) o sistemas SCADA tradicionales; además permiten muestrear las variables de interés con mayores resoluciones espaciales y temporales a lo alcanzable con métodos tradicionales.

El estándar IEEE 802.15.4:

El estándar IEEE 802.15.4 [7] define las características de la capa física y de la capa de control de acceso al medio (MAC) para redes inalámbricas de área personales (WPAN, Wireless Personal Area Networks) de baja tasa de transmisión. Las ventajas de utilizar el estándar IEEE 802.15.4 es que permite la utilización de dispositivos de fácil instalación que proveen transmisiones confiables a distancias cortas a un precio muy bajo, proporcionando un tiempo de vida razonable al utilizar fuentes de energía limitada (e.j. baterías alcalinas) y al mismo tiempo una pila de protocolos simple. Las características generales, presentadas, del estándar de comunicación IEEE 802.15.4, son: tasas de transferencias de 250Kb/S, 40Kb/S y 20Kb/S, manejo de redes en estrella y malla (peer-to-peer), direccionamiento corto con 16 bits y extendido con 64 bits, garantía del manejo de las ranuras de tiempo (GTS, guaranteed time slot), detección de los niveles de energía recibidos (ED, energy detection), indicadores de calidad en el enlace así como de conmutación de canales para recibir paquetes(LQI, link quality indication), acceso al canal por CSMA/CA (CCA, clear channel assessment).

Low-power and Lossy Networks (LLNs)

Las redes de bajo consumo y tolerantes a pérdidas (Low-power and Lossy Networks (LLNs)) pueden conectarse a un gran número de nodos, cuyos recursos (cómputo, energía) son limitados para formar una red inalámbrica ad-hoc. Los grupos de trabajo de la IETF como 6LoWPAN, ROLL [17] y CORE, han definido los protocolos para varias capas del stack de protocolos, incluyendo una adaptación de IPV6, un protocolo de ruteo y un protocolo de transferencia de datos web, que han sido basados en radios de bajo consumo IEEE 802.15.4. El modo *Timeslotted Channel Hopping* (TSCH) fue introducido en 2012 como un agregado a la MAC del estándar IEEE 802.15.4.

TSCH es un estándar emergente para la industria y el control de redes LLNs, inspirado en los ya existentes WirelessHART e ISA100.11a. Los nodos de una red IEEE802.15.4e [8] se pueden comunicar utilizando multiplexación por división de tiempo (TDMA). Una ranura de tiempo en un cronograma provee una unidad de ancho de banda que es reservada para la comunicación entre dos nodos vecinos. La reserva de la ranura de tiempo puede ser programada para que la comunicación sea predecible. Esto permite evitar que los nodos gasten energía en dejar sus trancptores en modo recepción a la espera de un paquete, extendiendo el tiempo de vida de las baterías de los dispositivos. Además la funcionalidad de salto de canales permite mayor confiabilidad ante la interferencia en un canal. 6TSCH [18] es una iniciativa del IETF cuyo objetivo es utilizar IPv6 sobre el modo de funcionamiento TSCH del estándar IEEE 802.15.4e, ya que IEEE 802.15.4e no define ningún método para construir y mantener el cronograma o sincronización entre nodos. Esta capa recibe el nombre de 6top [19] y ofrece interfaces de gestión y datos para capas superiores. Finalmente para la capa de aplicación podemos desarrollar protocolos propietarios, o basarnos en estándares abiertos, para lograr tener compatibilidad con otras aplicaciones, como REST (REpresentational State Transfer) [20] o COAP Constrained Application Protocol (CoAP) [21].

Especificación de IPv6:

La Internet de las Cosas debería poder acceder a entre 50 a 100.000 millones de objetos e incluso seguir sus movimiento o las variaciones de los parámetros sensados. Ésto es algo que no se puede hacer en la actualidad utilizando IPv4, el protocolo de ruteo de Internet, pues cada dispositivo necesita identificarse con una única numeración IP y como es de dominio público el direccionamiento Ipv4 [9] está agotado para asignar nuevas numeraciones. La IETF desde hace algunos años trabaja en un nuevo protocolo de Internet (protocolo IPv6), que permitió superar esta limitación [10]. Entre las mejoras que aporta respecto a IPv4, podemos enumerar las siguientes:

- Capacidad de direccionamiento expandida a 128 bits
- Calidad de servicio (Qos): IPv6 puede diferenciar los paquetes de datos como pertenecientes a un flujo particular, y así otorgar un ancho de banda en función de cada necesidad.
- Autoconfigurable (Neighbour Discovery): en IPv6 los nodos no necesitan ser configurados manualmente.

- End to end: IPv6 no usa NAT ya que tiene direcciones globales para todos los nodos.
- Simplificación del formato del encabezamiento:

Especificación 6LoWPAN

Contemplando las tendencias de la Internet de las Cosas y las redes de sensores inalámbricos de área personal de baja potencia, otro grupo de trabajo de la IETF desarrolló 6LoWPAN [11] que brinda soporte para redes de Sensores al protocolo IPv6.

La interacción de los nodos de sensores con IPv6 implica que cada red de sensores pueda fácilmente interactuar con cualquier otro nodo de cualquier red IPv6, incluyendo Internet. Lo que significa que cada sensor pueda ser accedido desde cualquier parte del mundo. El protocolo 6LoWPAN fue desarrollado para hacer posible el "Internet Embebido" simplificando las funcionalidades de protocolo de Internet IPv6, definiendo un encabezado muy compacto y tomando en cuenta la naturaleza de las redes inalámbricas.

El estándar define el formato de la trama de transmisión de paquetes IPv6, así como la configuración del direccionamiento IPv6 de link local y las direcciones globales configuradas automáticamente. Se define una capa de adaptación, debido a que IPv6 requiere el manejo de paquetes con tamaños mucho más grandes que el tamaño del tramas máximo de IEEE 802.15.4 (127 bytes). También se definen mecanismos para la compresión de encabezado IP, necesarios para hacer eficiente el uso de IPv6 sobre redes IEEE 802.15.4, y las disposiciones necesarias para la entrega de paquetes en redes de mallas [12].

Objetivos y Resultados

Objetivo Principal

Diseñar, simular, implementar y validar el comportamiento de una red de sensores inalámbricos de bajo consumo, a campo, y con conexión a IoT, en testbed, para la adquisición de parámetros ambientales agrícolas.

Objetivos secundarios

- Releva las exigencias de consumo de energía para el escenario de despliegue de la WSN en aplicaciones agrícolas.
- Desarrollar de una WSN eficiente energéticamente.
- Estudiar, simular y evaluar los protocolos de comunicación de internet de las cosas.
- Documentar y difundir los resultados.

Avances y resultados preliminares

Se desarrolló un prototipo de red basado en la norma IEEE 802.15.4 [7] para las capas física y enlace. Se diseñó e implementó un sistema de gestión que permite prever y determinar el tiempo de agotamiento de baterías, para un mote en particular. Para ello se usan como parámetros el hardware instalado, tipo de baterías y tipo de mote y el protocolo de administración de la energía utilizado.

Se trabajó en un sencillo algoritmo de enrutamiento [16] de datos específico para dicho prototipo, buscando maximizar el rendimiento energético del sistema, reduciendo el overhead. El método antedicho, permite minimizar el procesamiento necesario por parte del mote y reducir el consumo por la transmisión de paquetes pequeños.

Se implementó y ensayó un protocolo de propagación de sincronismo, determinando la mejor relación entre el bajo consumo y el menor error admisible. Se optó por un esquema centralizado, en el cual se resincroniza la red de manera pasiva, habiendo determinado de manera experimental la cantidad mínima de tiempo necesaria entre sincronizaciones sucesivas.

Se implementó un software de gestión en línea. Para el que se estudiaron casos de aplicación y se realizó un relevamiento de requerimientos para el desarrollo de la interfaz humana y de carácter funcional.

Se estudia y se realizaron las primeras pruebas simuladas de 6lowPAN para nodos IEEE802.15.4, utilizando S.O Contiki [13,15] y su simulador Cooja [15,14].

Formación de Recursos Humanos

Desde la creación del GridTICs, uno de los objetivos a realizar fue la capacitación de los recursos humanos tanto del grupo como externos. Esta actividad de formación se viene cumpliendo desde el comienzo, junto a la divulgación de las tecnologías de redes de sensores. En el punto 11, "Estado de avance" de esta misma presentación, se puede apreciar las actividades realizadas por el proyecto anterior para la formación de recursos humanos.

La meta como investigadores es fortalecer la capacidad para realizar investigación científica, generar conocimientos y facilitar la transferencia de tecnología que permita el desarrollo humano.

El camino para lograr este objetivo lleva a aumentar el número de científicos dedicados a la investigación en el país, incrementar la productividad científica e integrar a la comunidad

científica con el sector productivo y con el sector social. Por tanto es importante el desarrollo de infraestructura y el capital humano de ciencia y tecnología.

Este proyecto de investigación posibilita la colaboración inter-institucional y la ejecución de proyectos conjunto entre grupos I+D de diferentes disciplinas entre las cuales podemos mencionar:

- Agricultura de precisión, Software embebido, Calidad de software, Redes de datos, sensado, mediciones de parámetros.

Para lograr estos objetivos se desarrollarán:

- Dictado de Cursos, Seminarios y Conferencia para público especializado de la región.

- Promoción, coordinación y asistencia técnica de tesis de grado para alumnos de Ingeniería de Sistemas de Información e Ingeniería Electrónica de la FRMza

- Promoción, coordinación y asistencia técnica a pasantes alumnos, cursantes de carreras de grado y de pre-grado (Academia Cisco, tecnicaturas, etc) en el ámbito de la UTN FRMza

- Promoción, coordinación, dirección y asistencia técnica a Tesis doctorales, postgrado y/o maestría.

- Presentación de Trabajos en Congresos y Reuniones Técnicas/Científicas.

- Publicación de Trabajos en revistas con/sin referato.

Referencias

[1] Verdtech S.A., website:
<http://www.verdtech.es/>

[2] Adcon Telemetry, website
<http://www.adcon.com/index.php?lang=en>

[3] Crossbow's technology, <http://www.moog-crossbow.com/>

[4] Universidad de Colima y Universidad de Valencia. <http://www.d2ars.org>

[5] Libelium, website <http://www.libelium.com/>

[6] "Wireless Sensor Networks" Ian Akyildiz, Mehmet Can Vuran - Wiley - 2010.

[7] IEEE Std 802.15.4-2003: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)," IEEE Standard, 2003

[8] 802.15.4e-2012 - IEEE Standard for Local and metropolitan area networks--Part 15.4: Low-Rate

Wireless Personal Area Networks (LR-WPANs) Amendment 1: MAC sublayer
<http://standards.ieee.org/getieee802/download/802.15.4e-2012.pdf>

[9] Informe LACNIC, "Distribuciones/Asignaciones IPv4, espacio disponible y pronósticos (Informe Abril 2012)", <http://www.lacnic.net/sp/registro/espacio-disponible-ipv4.html>

[10] Robert L. Fink "IPv6—What and Where It Is", The Internet Protocol Journal, Volume 2, Number 1, March 1999

[11] N. Kushalnagar, G. Montenegro, and C. Shumacher, "IPv6 over lowpower wireless personal Area networks (6lowpans)," RFC 4919, Aug. 2004.

[12] G. Montenegro, N. Kushalnagar, J. Hui, and D. Culler, "Transmission of IPv6 packets over ieee802.15.4 networks," RFC 4944, Sep. 2007.

[13] Adam Dunkels, Bjorn Gronvall, and Thiemo Voigt. Contiki-a lightweight and exible operating system for tiny networked sensors. In Local Computer Networks, 2004. 29th Annual IEEE International Conference on , pages 455-462. IEEE, 2004

[14] Osterlind, Fredrik, et al. "Cross-level sensor network simulation with cooja." *Local computer networks, proceedings 2006 31st IEEE conference on*. IEEE, 2006.

[15] Contiki Os & cooja simulator website:
<http://www.contiki-os.org/>

[16] Diedrichs, Ana Laura, Germán Tabacchi, Guillermo Grunwaldt, Matías Pecchia, Gustavo Mercado, and Francisco Gonzalez Antivilo. "Low-power wireless sensor network for frost monitoring in agriculture research." In *Biennial Congress of Argentina (ARGENCON), 2014 IEEE*, pp. 525-530. IEEE, 2014.

[17] RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks -
<http://tools.ietf.org/html/rfc6550>

[18] 6tsch architecture
<http://tools.ietf.org/html/draft-thubert-6tsch-architecture-02>

[19] 6top <http://tools.ietf.org/html/draft-wang-6tsch-6top-00>

[20] Richarson, Leonard ; Amundsen, Mike "RESTful Web APIs" - O'Reilly - 2013

[21] Constrained Application Protocol (CoAP)
<https://datatracker.ietf.org/doc/draft-ietf-core-coap/>

GW-CIAA-IoT: Gateway con CIAA para red inalámbrica de IoT

Carlos Taffernaberry, Gustavo Mercado

GridTICs – Grupo en Tecnologías de la Información y las Comunicaciones /
Departamento de Electrónica / Facultad Regional Mendoza / Universidad
Tecnológica Nacional

Rodríguez 273 Capital – Mendoza, +54 261 5244563
{gustavo.mercado, carlos.taffernaberry}@gridtics.frm.utn.edu.ar

Resumen

La comunicación autónoma entre objetos sin la participación de un ser humano se conoce hoy como la Internet de las Cosas (IoT). Estas *cosas* son dispositivos electrónicos embebidos dentro de objetos de uso cotidianos permitiendo su conexión a Internet para lograr el acceso remoto a los datos generados, o el accionamiento de actuadores a distancia.

Son implementados generalmente con microcontroladores con bajo poder de procesamiento, poseen comunicación inalámbrica y sensores, permitiendo obtener valores de parámetros del medio ambiente tales como temperatura, humedad, posición, etc. Estos objetos normalmente pueden interactuar con cualquier otro dispositivo conectado a Internet.

El presente proyecto pretende evaluar los distintos protocolos de comunicación de la Internet de las Cosas y aplicarlos a una red de sensores existente, llamada SIPIA implementada en el ámbito del sector agrícola como la agricultura de precisión.

Para llevar a cabo la implementación se utilizará la Computadora Industrial Abierta Argentina (CIAA), que es una plataforma electrónica libre y gratuita

preparada especialmente para trabajar en aplicaciones industriales. Fue desarrollada por medio del trabajo colaborativo, transdisciplinario entre instituciones, academias y empresas de la Argentina .

Palabras clave: Internet de las Cosas, Red SIPIA, Redes de sensores, CIAA, Computadora Industrial Abierta Argentina, 6lowPAN, IEEE 802.4.15,

Contexto

Este trabajo está inserto en el proyecto de investigación y desarrollo tecnológico denominado "RED SIPIA-LP: Estudio de mecanismos de bajo consumo energético para aplicar a una red de sensores inalámbricos en el ámbito de agricultura de precisión", homologado y financiado con el código UTI3646TC por Rectorado de la Universidad Tecnológica Nacional [1].

El Ministerio de Ciencia, Tecnología e Innovación Productiva, a través de la Subsecretaría de Políticas en Ciencia, Tecnología e Innovación Productiva, convocó a la presentación de propuestas destinadas a la adjudicación de "Proyectos de Innovación a partir de la adopción de la Computadora Industrial Abierta Argentina en productos y procesos industriales". El presente trabajo fue presentado, aprobado y financiado

por el Ministerio de Ciencia, Tecnología e Innovación Productiva, de acuerdo a la resolución 613/2015 [2]

Introducción

Gracias a los avances y reducción de costos en dispositivos electrónicos y de comunicación inalámbrica, es posible en la actualidad construir dispositivos sensores multifuncionales y multipropósito de bajo costo que operan con poca energía, pequeño tamaño, y con capacidad de comunicación a corta distancia. Estos dispositivos, denominados comúnmente motes, constan de una unidad de procesamiento con un poder de cómputo mínimo, memoria, un módulo de comunicación inalámbrica y uno o varios dispositivos de sensado que capturan parámetros como temperatura, aceleración, humedad, etc. Un conjunto de motes comunicados entre sí es lo que se conoce como una red de sensores inalámbrica (Wireless Sensor Network o WSN) [3].

Debido a las WSN, en los últimos años, han surgido nuevas tendencias en el sector agrícola como la agricultura de precisión. Esta disciplina cubre múltiples prácticas relativas a la distinta toma de datos para aplicar en la gestión y toma de decisiones de cultivos y animales. Por medio de sensores estratégicamente situados, se puede realizar un monitoreo permanente de parámetros ambientales como la temperatura y/o humedad relativa, con el fin de proveer de gran cantidad de datos confiables para la investigación agronómica.

En ese sentido el grupo GridTICS en conjunto con UNCuyo, desarrollo el proyecto SIPIA (proyecto homologado Código C&T UTN UTI1737) que permitió desarrollar un prototipo de red basado en la norma IEEE 802.15.4 [4]

para las capas física y de enlace. Para las capas superiores en lugar de utilizar estándares, se prefirió realizar un desarrollo propio, tomando como prioridad la optimización de la energía, debido a que los motes sensores fueron alimentados con baterías. No se diseñó algoritmo alguno de enrutamiento de datos, pues la topología que se adoptó fue estrella, siendo el centro de la misma un mote coordinador. Se implementó y ensayó un protocolo de propagación de sincronismo, determinando la mejor relación entre el bajo consumo y el menor error admisible. Se optó por un esquema centralizado, en el cual se resincronizaba la red de manera pasiva, habiendo determinado de manera experimental la cantidad mínima de tiempo necesaria entre sincronizaciones sucesivas.

Se diseñó e implementó un sistema de gestión que permitió obtener periódicamente los valores sensados, y también prever y determinar el tiempo de agotamiento de baterías para cada mote en particular. El software fue instalado en una computadora ubicada en la red local en la Facultad de Ciencias Agrarias (UNCuyo), cercana al viñedo donde se ubicó la WSN. Como vinculación entre ambos tipos de redes se debió diseñar y construir un dispositivo que denominamos Gateway.

El Gateway debió conectarse a ambas redes, por un lado a la WSN, limitando su ubicación a un área del viñedo próxima a la WSN, debido al escaso alcance de los enlaces de la norma IEEE 802.15.4 usada. Por otro lado, se conectó a la red local utilizando un enlace WiFi (IEEE 802.11); debido a la exigencia de energía de este último protocolo, el Gateway debió ser alimentado por medio de la red eléctrica.

El software embebido en el Gateway permitió que los datos provenientes de la

red de sensores, sean interpretados, convertidos, almacenados y luego transmitidos hasta la computadora que alberga el software de gestión.

Por otro lado, hace 20 años, Internet se usaba principalmente como herramienta para acceder a información. En los últimos 10 años hemos vivido una nueva forma de uso de Internet, donde todo se ha convertido en social, transaccional y móvil. En la actualidad estamos sufriendo una nueva transformación, en la que cada objeto tiene una identidad virtual propia y es capaz de integrarse e interactuar de manera independiente con cualquier otro objeto, sistema, o humanos. De esta forma estamos en presencia de un nuevo cambio en nuestra forma de vida, creándose nuevos modelos de negocio, productos y compañías, denominándose a esto la Internet de las Cosas (IoT) [5]. Este desarrollo no sería posible sin el soporte del nuevo protocolo de red de internet, llamado IPv6 [6], que permite contar con direcciones de red suficientes como para dotar a cada uno de los componentes de una red de sensores de una dirección pública IPv6.

Puntualmente, el ámbito de las WSN ha despertado un interés especial para lograr asociar a cada uno de los motes que la componen a la IoT. Contemplando los requerimientos de la IoT y las WSNs de área personal (PAN), un grupo de trabajo de la IETF desarrolló el protocolo 6LoWPAN [7] que brinda soporte a redes LowPAN para el protocolo IPv6.

Por medio de este protocolo, como se puede observar en el Gráfico 1, 6LoWPAN habilita la interacción de los nodos de sensores con el backbone de Internet, permitiendo una comunicación de pares (p-2-p) de cada sensor y cualquier host en cualquier lugar del mundo.

Para permitir extender el alcance de una WSN, debido a que no siempre todos los nodos tienen alcance entre sí, se creó otro grupo de trabajo de la Task Force de la IETF llamado ROLL (Routing Over Low power and Lossy Networks) para evaluar problemas de encaminamiento dentro de dichas redes y proponer soluciones.

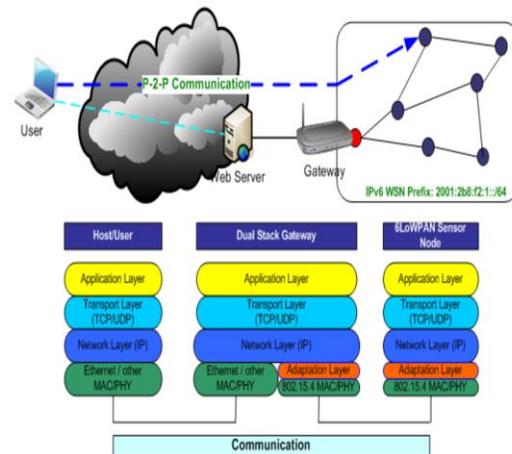


Gráfico 1

El objetivo del grupo de trabajo fue el diseño de un protocolo de enrutamiento para LLN (Low power and Lossy Networks) llamado RPL [8], con soporte de una variedad de capas de enlace (IEEE 802.15.4, IEEE 802.15.4g o Powerline Communication), que comparten características comunes, como ser bajo ancho de banda, alta tasa de pérdidas y baja potencia.

De lo expuesto anteriormente, se puede observar que el proyecto SIPIA tiene algunas limitaciones tecnológicas, detalladas a continuación:

- Limitación referida a la máxima extensión territorial de la WSN: estará limitada por el área de cobertura de las transmisiones IEEE 802.15.4, por no implementar protocolos de ruteo entre los motes, dentro de la misma red.

- Limitación de conversión de protocolos: es debido al uso de tecnologías de capa de

red distintas en ambos tipos de redes. Por un lado un protocolo propietario en la WSN, y por otro lado protocolo IPv4 en la red Lan.

- No existe una comunicación end-to-end entre la aplicación de gestión y la aplicación que se ejecuta en los motes.

Estos y otros problemas son los que motivaron a convertir la red de sensores SIPIA a una red que utilice los protocolos de la Internet de las Cosas, permitiendo sortear las limitaciones previamente expuestas.

Para ello, se propone utilizar la CIAA [9] para diseñar e implementar un nuevo Gateway entre ambas redes, y un reemplazo de los motes por nuevos que soporten los protocolos de IoT.

Líneas de investigación y desarrollo

El grupo de Redes de Sensores Inalámbricos, perteneciente al grupo UTN GridTICS, se forma por iniciativa de algunos de sus miembros en el año 2008 y comienza con el estudio de la tecnología para la capacitación y luego la adquisición de elementos y dispositivos para la conformación del laboratorio de WSN. El grupo ha realizado publicaciones, presentación en congresos, cursos de grado y posgrado y asistencia a tesinas de grado y tesis de posgrado en curso.

El grupo de tecnología IPv6, perteneciente al grupo UTN GridTICS, se constituye en 2005 y ha tenido una vasta actividad y experiencia y es reconocido como uno de los grupos pioneros en IPv6 de la región. El grupo ha realizado publicaciones, presentación en congresos, cursos de grado y posgrado y asistencia a tesinas de grado y tesis de posgrado. Además de participar activamente en las

iniciativas de ISOC y de LANIC para la promoción y difusión de IPv6, siendo también socio activo de la IPv6 Task Force Argentina.

Resultados y Objetivos

El objetivo principal es:

- Diseñar, ensayar e implementar una mejora de un sistema de redes inalámbricas agrícolas aplicadas, usando tecnología CIAA.

Mientras que se detallan los objetivos específicos a continuación:

- Análisis y estudio de protocolos de comunicación de la IoT aplicado a WSN.

- Diseño y desarrollo de hardware de Gateway de red SIPIA de WSN utilizando CIAA.

- Evaluación y selección de sistema operativo embebido aplicable a CIAA con soporte de protocolos IoT.

- Diseño y desarrollo de software de Gateway de red SIPIA de WSN utilizando CIAA.

- Evaluación, selección e Implementación de motes de WSN con soporte de IoT.

- Aportar al proyecto libre y colaborativo CIAA de las siguientes tecnologías:

- Hardware y software específico de protocolo IEEE 802.15.4 de WSN
- Implementación de Software específico de protocolo IETF 6LowPan de WSN
- Implementación de Software específico de protocolo IETF RPL de WSN

- Capacitación y difusión de los conocimientos adquiridos, como compromiso social en el ámbito empresarial y académico.

Formación de Recursos Humanos

Desde la creación del GridTICs, uno de los objetivos a realizar fue la capacitación de los recursos humanos tanto del grupo como externos. Esta actividad de formación se viene cumpliendo desde el comienzo, junto a la divulgación de las tecnologías de redes de sensores.

La meta como investigadores es fortalecer la capacidad para realizar investigación científica, generar conocimientos y facilitar la transferencia de tecnología que permita el desarrollo humano.

El camino para lograr este objetivo lleva a aumentar el número de científicos dedicados a la investigación en el país, incrementar la productividad científica e integrar a la comunidad científica con el sector productivo y con el sector social. Por tanto es importante el desarrollo de infraestructura y el capital humano de ciencia y tecnología.

Para lograr estos objetivos se desarrollarán:

- Dictado de Cursos, Seminarios y Conferencia para público especializado de la región.

- Promoción, coordinación y asistencia técnica de tesis de grado para alumnos de Ingeniería de Sistemas de Información e Ingeniería Electrónica de la FRMza

- Promoción, coordinación y asistencia técnica a pasantes alumnos, cursantes de carreras de grado y de pre-grado en el ámbito de la UTN FRMza

- Promoción, coordinación, dirección y asistencia técnica a Tesis doctorales, postgrado y/o maestría.

- Presentación de Trabajos en Congresos y Reuniones Técnicas/Científicas.

- Publicación de Trabajos en revistas con/sin referato.

Referencias

[1] <http://sicyt.scyt.rec.utn.edu.ar/sicyt/proyectos/RESUMENF.ASP?VAR1=3646>

[2] <http://www.mincyt.gob.ar/adjuntos/archivos/000/042/0000042058.pdf>

[3] Akyildiz, Ian F., and Mehmet Can Vuran. Wireless sensor networks. Vol. 4. John Wiley & Sons, 2010.

[4] MOLISCH, Andreas F., et al. IEEE 802.15. 4a channel model-final report. IEEE P802, 2004, vol. 15, no 04, p. 0662.

[5] CONNER, Margery. Sensors empower the "Internet of Things". EDN (Electrical Design News), 2010, vol. 55, no 10, p. 32.

[6] HUITEMA, Christian. *IPv6: the new Internet protocol*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1996.

[7] KUSHALNAGAR, Nandakishore, et al. IPv6 over low-power wireless personal area networks (6LoWPANs): overview, assumptions, problem statement, and goals. RFC 4919 (Informational), Internet Engineering Task Force, 2007.

[8] VASSEUR, J., et al. RPL: The IP routing protocol designed for low power and lossy networks. Internet Protocol for Smart Objects (IPSO) Alliance, 2011, vol. 36.

[9] <http://www.proyecto-ciaa.com.ar/>

PROLONGANDO VIDA UTIL DE REDES DE SENSORES. MODIFICACION PROTOCOLO AODV

Juan de Dios Benitez, Ricardo Urdinola, Micaela Olivera, Diego Alberto Godoy,
Eduardo O. Sosa, Edgardo A. Belloni, Hernán Bareiro

Centro de Investigación en Tecnologías de la Información y Comunicaciones (C.I.T.I.C.) -
Universidad Gastón Dachary (UGD)

{juan.benitez, rurdinola, molivera, diegodoy, eduardo.sosa, ebelloni, hbareiro} @citic.ugd.edu.ar,

Resumen

En el presente trabajo se presenta una modificación en el protocolo de ruteo multi salto reactivo AODV para WSN el cual permite por medio de la adaptación del calculo de su métrica, poder aumentar la vida de una red disminuyendo la probabilidad de la existencia de segmentaciones en una WSN. Permitiendo de esta manera la optimización de la batería de los nodos extendiendo la vida útil de nodo individual, concluyendo con una vida útil de la red mejorada y aumentada. Dicho estudio es una línea dentro del proyecto denominado “*Diseño de arquitecturas de soporte a la Internet del futuro y Ambientes Inteligentes para su Aplicación en Ciudades Inteligentes – Fase II*” de la Universidad Gastón Dachary.

Palabras clave: Protocolo, AODV, Protocolo Reactivo, Redes de Sensores Inalámbricos, IoT

Contexto

El trabajo se enmarca en el proyecto de investigación denominado “*Diseño de Arquitecturas de Soporte a la Internet del Futuro y Ambientes Inteligentes*” y “*Diseño de arquitecturas de soporte a la Internet del futuro y Ambientes Inteligentes para su Aplicación en Ciudades Inteligentes – Fase II*”, acreditados en la Secretaría de Investigación y Desarrollo de la UGD por RR.19/A/12 y RR.18/A/14 respectivamente.

Se relaciona y articula directamente con diversos proyectos de investigación acreditados en la Secretaría de Investigación y Posgrado (SECIP) de la Facultad de Ciencias Exactas, Químicas y Naturales de la UNaM, enfocados en las temáticas “*Internet del Futuro*” e “*Internet de las Cosas*”, entre los que se incluyen: Proyecto 16Q457 “*Hacia la Programación de Sensores Inalámbricos en la Forma Web 2.0*”; Proyecto 16Q474 “*Simulaciones de Sistemas Modernos de Comunicación*”; y Proyecto 16Q519 “*Ambientes Inteligentes. Una Mirada a Internet del Futuro*”.

Referente a este proyecto existen 5 tesinas de grado en curso, 3 tesinas de grado finalizadas, un trabajo final de Especialización finalizado, un trabajo de fin de Máster en curso y una Tesis Doctoral en etapa de redacción final.

Introducción

El aumento de la capacidad de cómputo, la miniaturización de los componentes y el abaratamiento de los dispositivos hacen posible la realización de la Internet de las Cosas (IoT, Internet of Things) [1] y de la Inteligencia Ambiental [2], en la que los objetos cotidianos contarán con sensores conectados a Internet.

El desarrollo de estas tecnologías brinda oportunidades de aprovechamiento económico e industrial, y a la vez representa un enorme desafío tecnológico. Como parte del desarrollo de las tecnologías de la IoT, las Redes de Sensores Inalámbricos (WSN, Wireless Sensor Network) [3] son un área de continuo estudio y desarrollo [4]. Hoy en día

existen WSNs con miles de nodos, recolectando datos en infinidad de entornos, para todo tipo de fines y a un costo razonable.

Objetivos

El objetivo general del proyecto marco es: "Diseñar arquitecturas de soporte a Internet del futuro y Ambientes Inteligentes para su aplicación a Ciudades Inteligentes". Como objetivos específicos se indican: A) Especificar el Estado del Arte de la Internet del de Futuro y los Ambientes y Ciudades Inteligentes. B) Determinar los protocolos de ruteo óptimos para WSN y ciudades inteligentes. C) Determinar el impacto en la vida útil de una red WSN de del protocolo AODV estándar por medio de simulaciones. D) Analizar y diseñar una modificación a dicho protocolo para que el mismo sea energéticamente eficiente aumentando la vida útil de una red. E) Simular escenarios de Ciudades inteligentes de redes de sensores inalámbricos. F) Determinar cuál es la métrica correcta que optimiza la vida útil de la red. G) Realizar las simulaciones entre los protocolos AODV estándar y AODV modificado, para analizar la mejora en la red.

Protocolo AODV

Una de las características más importantes que poseen los nodos en redes WSN es la habilidad de crear redes autónomas. Esto hace realmente fácil la implementación y el mantenimiento de las WSN, pero requiere que los nodos implementen dos funciones muy diferentes: La captura de información del medio para la cual fue distribuido y a su vez el encaminamiento de los paquetes desde el origen hasta el destino seleccionado.

Generalmente, cuando una WSN es implementada, existe un nodo denominado Gateway o nodo sumidero, que posee mayores recursos, tanto energéticos como de hardware, que los otros nodos. Este nodo sink (sumidero) suele tener acceso a internet (si fuera necesario) y fuentes de energía ininterrumpidas. Por otro lado los demás nodos que conforman la red son realmente limitados en lo que se refiere a,

almacenamiento, potencia de cómputo y esencialmente a la capacidad de energía la cual es provista por baterías internas.

Esta última limitación es un factor crítico a considerar. El tiempo de vida de la red depende directamente de la energía residual con la que cuenta cada nodo. Se entiende por energía residual a la energía remanente, que poseen las baterías al ser utilizadas. A medida que los nodos se quedan sin con poca energía, la red puede llegar a sufrir fragmentaciones, causando lo que se denominan islas, las cuales son inalcanzables por el resto de la red a raíz de la limitada potencia de transmisión que poseen los nodos.

La complejidad se incrementa en las redes que implementan métodos de reenvío de paquetes por medio de multisaltos, donde múltiples retransmisiones son necesarias para poder enviar un paquete desde el origen hasta el destino. El mayor consumo de energía se lleva a cabo en los momentos de Transmisión y recepción de información por medio de los nodos. Esto implica que los protocolos y algoritmos de ruteo son claramente esenciales y decisivos en el tiempo de vida de una WSN. Sin embargo uno de los algoritmos mas utilizados por redes WSN como es el AODV (On- Demand Distance Vector Routing) – Un protocolo de camino más corto, sigue un enfoque tradicional con una búsqueda del camino con el costo más bajo a través del conteo de saltos, causando que los nodos participantes en la red concentren su tráfico a través de un grupo de nodos aumentando el consumo de energía en los mismos y de esta forma segmentando la red de manera más rápida. El protocolo *AODV*, es un protocolo de ruteo proactivo el cual actúa únicamente cuando hay la necesidad de transmitir a un destino en particular a través de una ruta desconocida. El proceso se activa iniciando el descubrimiento del camino, lo que básicamente se lleva a cabo por medio de una inundación de la red a través de un paquete de *RREQ* (*Route Request*). Cuando el paquete *RREQ* llega al nodo deseado, un paquete de *RREP* (*Route Replay*) es enviado al nodo origen de por medio de una transmisión en unicast como se puede apreciar en la Fig 1.

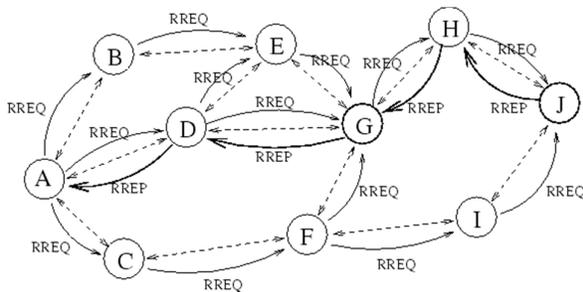


Fig 1: Proceso de descubrimiento de la ruta

El paquete *RREQ* (Fig 2) cuenta con un campo de 8 bits denominado “Hop Count” (conteo de saltos), el cual es incrementado en cada retransmisión del paquete por todos los nodos participantes.

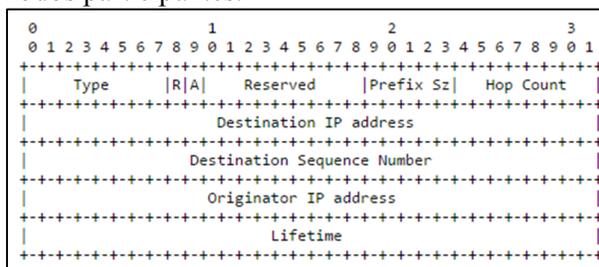


Fig 2: Descripción del paquete RREQ

Cuando el paquete *RREQ* llega al destino, el campo de conteo de saltos tiene el valor necesario para transmitir el paquete entre el nodo origen y el destino. Esta métrica de saltos, representa la distancia física entre ambos nodos. Este costo es almacenado en el nodo, en su tabla de ruteo y determina el camino que se debe utilizar para alcanzar dicho destino

Modificación del Protocolo AODV

La modificación que se ha realizado en el algoritmo de cálculo de la métrica del protocolo AODV estándar para mejorar la vida de la red, consiste en que durante el proceso de descubrimiento, los nodos al incrementar el costo en el campo de *hop count* en el *RREQ*, en vez de incrementar el valor en una unidad por cada salto, adicionará la energía residual que tiene el nodo que retransmite el paquete en sus baterías, por medio de una aproximación cuadrática. La cual permite una priorización de la ruta mas adecuada frente a otros tipos de aproximaciones como ser lineales,

exponenciales negativas, etc. Estos parámetros se han concluido a través de análisis empíricos de los distintos cálculos de métricas y simulaciones de desempeño de la red.

La ecuación propuesta para determinar este costo es:

$$Cost = -\left(102 + \frac{8}{21}\right)x^2 - \left(152 + \frac{13}{21}\right)x + 255$$

Donde “*x*” es la energía residual relativa en el nodo. Esta ecuación representa una función cuadrática, obteniendo un costo de 255 cuando el nodo no tiene energía y un costo de 0 cuando la energía residual relativa es igual a 1 (Fig 3).

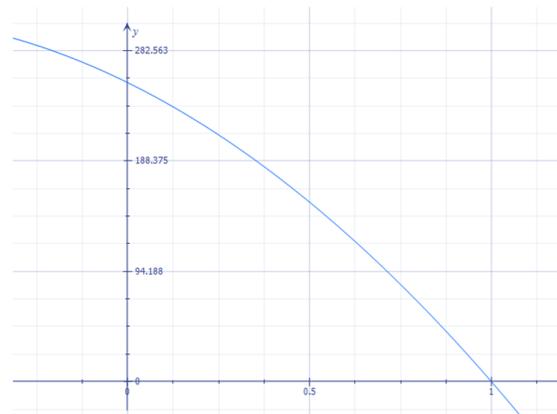


Fig. 3 Función Cuadrática

Proceso de Simulación

El entorno usado para realizar los análisis y simulaciones cuenta con una PC estándar con un sistema operativo Linux distribución Ubuntu 14.04, en el cual se corre el simulador Omnet++, un simulador de eventos discretos orientado a objetos, distribuido bajo licencia publica académica. Una de las características más importantes del simulador proviene de la capacidad del uso adicional de *frameworks*, los cuales habilitan a la implementación sencilla de los módulos necesarios.

El equipo de desarrollo de Omnet++, mantiene los *frameworks* llamados INET. Sin embargo para este trabajo se ha utilizado una extensión denominada INETMANET, la cual provee la implementación de MANETS, incluyendo el protocolo AODV, desarrollado

por la Universidad Upsala. Esta Extensión ha sido usada y testeada por la comunidad Omnet++ por alrededor de 10 años, por lo que se considera una extensión estable.

Se han llevado a cabo dos series de simulaciones. en la primera de ellas se realizó utilizando el protocolo AODV estándar, mientras que en la segunda fue realizada bajo las mismas características que la anterior pero utilizando el protocolo AODV modificado que se presenta en este trabajo. Cada serie cuenta con 50 corridas del simulador configuradas con diferentes semillas en los generadores números aleatorios, produciendo escenarios distintos. Para poder mantener la coherencia entre las series de simulaciones, se configura al simulador para mantener el mismo juego de semillas entre las dos series, de esta forma se pueden comparar ambos protocolos bajo las mismas condiciones de aleatoriedad.

Cada corrida del simulador consiste en la generación aleatoria de un escenario de 1Km² el cual cuenta con 25 nodos sensores distribuidos de manera aleatoria con una probabilidad uniforme. La densidad de nodos resultante permite una alta probabilidad del uso de ruteo multisalto, para realizar el encaminamiento de los paquetes desde un nodo origen hasta el nodo *gateway* o *sink*.

Como se desea que la simulación represente un escenario lo más real posible, se han configurado los nodos de manera que tengan ciclos de transmisión de datos aleatorios en un intervalo entre 50 y 120 segundos.

Todos los nodos están equipados con un modelo de consumo de batería lineal, implementado en el *framework* INETMANET, y el consumo de energía es el único factor determinante en la muerte de los nodos. Los nodos se encuentran equipados con baterías de 3800mAh con un voltaje de 12V.

En la Tabla 1, se puede observar el consumo de energía original configurado en la simulación.

Tabla 1 Consumo energía original.

<i>Estado del nodo</i>	<i>Valor</i>
Ocupado	0.5mA
Recepción	1mA
Espera	1mA
Dormido	0.001mA

Para poder reducir los tiempos de simulación, y debido a las limitaciones de hardware para las simulaciones se ha decidido incrementar cada uno de los consumos en diez mil veces, dando como resultado los valores que se puede ver en la Tabla 2:

Tabla 2 Consumo de Energía Incrementado

<i>Estado del nodo</i>	<i>Valor</i>
Ocupado	5000mA
Recepción	10000mA
Espera	10000mA
Dormido	10mA

Como el interés de este trabajo es la medición del tiempo de vida de la red, se ha decidido que cada simulación se corte al momento que el primer nodo se queda sin batería y muere. En ese momento se realiza la medición de la energía residual de la red para comparar con el protocolo original.

Resultados

Usando los resultados de las simulaciones, se han comparado los parámetros de tiempo de vida y distribución de energía en la red utilizando ambos protocolos, AODV estándar y AODV modificado.

Como se puede apreciar en la Fig 4, el protocolo AODV modificado tiene un tiempo de vida de la red mayor que el AODV estandar.

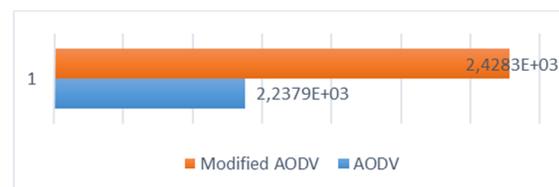


Fig. 4 Comparación del tiempo de vida promedio [Segundos]

Tabla 3 Mediciones de Tiempo de Vida promedio de la red

	<i>AODV</i>	<i>Modif. AODV</i>	<i>Diferencia</i>
Tiempo	2237,9338s	2428,2961s	8,5062%

El resultado muestra una mejora del 8,5% en la utilización del AODV modificado contra el AODV estándar.

Con la utilización del protocolo AODV modificado, se ha buscado obtener una mejor distribución del consumo de energía en la red. El resultado esperado se muestra a través del uso de la desviación estándar utilizada en la representación de una distribución normal.

En la Fig 5 se puede observar los resultados de las simulaciones en notación científica.

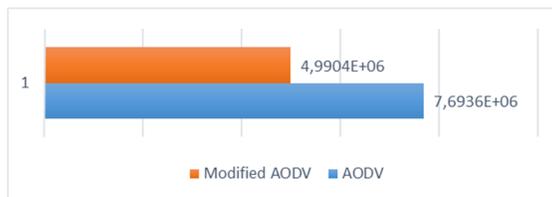


Fig. 5 Comparación de la Desviación Estándar

Como se puede ver en la Fig 5 se ha logrado una reducción del 35,13% en la desviación estándar del protocolo modificado sobre el protocolo estándar. Esto significa que los nodos distribuyeron su energía de una manera mas homogénea en la red.

En la practica esto es equivalente a decir que los paquetes utilizaron diferentes rutas, evitando que se concentre una ruta sobre un mismo nodo agotando de manera mas rápida su energía.

Este resultado se puede observar mejor en el estudio de la distribución normal, donde el protocolo AODV con la métrica modificada es más estrecho (Fig 6).

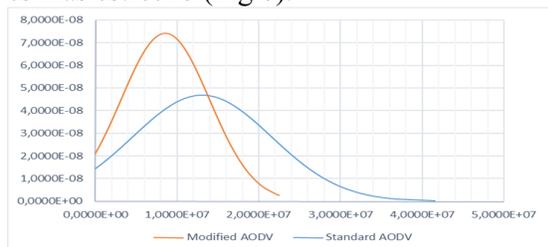


Fig. 6 Comparación de la distribución Normal

De otra manera, se puede observar que el protocolo modificado, tiene una curva mas hacia la izquierda que el AODV estándar, con lo que se puede deducir que la curva del protocolo modificado posee una energía residual media menor que el protocolo estándar.

Este es el efecto esperado y deseado en la comparación de ambos protocolos.

Formación de Recursos Humanos

El equipo de trabajo se encuentra formado por una Doctora en Ingeniería Telemática(España), un Doctor en Ciencias Informáticas, un Doctorando en Tecnologías de la Información y Comunicaciones, Magister y Especialista en Ingeniería de Software, un Maestrando de Ingeniería de la Web, un Maestrando de Redes de Datos, tres auxiliares de investigación graduado, un auxiliar de investigación graduado residente en la ciudad de Córdoba Capital, y ocho auxiliares de investigación en período de realización de trabajos de grado. El número de tesis de grado en curso con proyecto aprobado es de cinco y el número de trabajos de especialidad finalizado es tres dentro de la línea de investigación. Los proyectos de grado se titulan “Diseño de un prototipo para monitoreo eficiente de iluminación basado en WSN utilizando HTML5”, “Contribución a la Gestión de Residuos Domiciliarios como una Aplicación en Ciudades Inteligentes” y “Análisis y comparación de modelos de propagación para optimizar la localización geográfica de Ganado”.

Bibliografía

- [1] K Ashton, "“That 'Internet of Things' Thing”," 2009 (rev. 2011).
- [2] Ahola J., Ambient Intelligence, 2001.
- [3] W. Dargie and C. Poellabauer, “Fundamentals of Wireless Sensor Networks - Theory and Practice. Reino Unido: Wiley, West Sussex, 2010.
- [4] Eduardo O. Sosa, Contribuciones al establecimiento de una red global de Sensores Inalámbricos. Tesis Doctoral, Junio 17, 2011.
- [5] DerWerff, T. J. Van, «10 emerging technologies that will change the world,» MIT

- Technology Review, vol. 2, 2003.
- [6] R. M. Perez, «Algoritmos Fiables y Eficientes basados en Enrutamiento Geográfico para Redes Realistas de Sensores Inalámbricos,» Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, Murcia, 2012.
- [7] «The Alliance,» ZigBee Alliance, 2014. [En línea]. Available: <http://www.zigbee.org/About/AboutAlliance/TheAlliance.aspx>. [Último acceso: 11 06 2014].
- [8] «ZigBee Specification Overview,» ZigBee Alliance, 2014. [En línea]. Available: <http://www.zigbee.org/Specifications/ZigBee/Overview.aspx>. [Último acceso: 11 06 2014].
- [9] J. P. Dignani, *Análisis del Protocolo Zigbee*, La Plata, Buenos Aires: Facultad de Informática. Universidad Nacional de La Plata, 2011.
- [10] D. Gascón, «802.15.4 vs ZigBee,» Wireless Sensor Networks Research Group, 17 11 2008. [En línea]. Available: <https://web.archive.org/web/20150206055338/http://www.sensor-networks.org/index.php?page=0823123150>. [Último acceso: 11 06 1014].
- [11] Huarui Wu, Chunjiang Zhao, Li Zhu, «Study on an Energy-aware Routing Algorithm for Agriculture WSN,» *TELKOMNIKA*, vol. 11, nº 7, p. 3576 ~ 3584, 2013.
- [12] Fatma Bouabdallah, Nizar Bouabdallah, Raouf Boutaba, «On Balancing Energy Consumption in Wireless Sensor Networks,» Institut National de Recherche en Informatique et en Automatique, School of Computer Science, University of Waterloo, Rennes, France; Waterloo, Canada, 2008.
- [13] Ruben Hidalgo, José Ignacio Moreno, «Routing Design in Wireless Sensor Networks and a Solution for Healthcare Environments,» *IEEE LATIN AMERICA TRANSACTIONS*, vol. 9, nº 3, 2011.
- [14] G. Werner-Allen, K. Lorincz, M. Welsh, O. Marcillo, J. Johnson, M. Ruiz, «Deploying a wireless sensor network on an active volcano,» *IEEE Internet Computing*, vol. 10, no. 2, pp. 18–25, 2006.
- [15] Q. Wang, M. Hempstead, and W. Yang, «A realistic power consumption,» de *Sensor and Ad Hoc Communications and Networks*, 3rd Annual IEEE Communications Society on, vol. 1, 2006, pp. 286-295.
- [16] Diego Martínez, Francisco Blanes, José Simo, Alfonso Crespo, *Redes de Sensores y Actuadores Inalámbricas: Una Caracterización y Caso de Estudio para Aplicaciones Médicas en Espacios Cerrados*, Universidad Rovira i Virgili de Tarragona, España. : XXIX Jornadas de Automática. , 2008.
- [17] Héctor Ramos Morillo, Francisco Maciá Pérez, Diego Marcos Jorquera , *Redes Inalámbricas de Sensores Inteligentes. Aplicación a la Monitorización de Variables Fisiológicas*, Alicante, España : Departamento de Tecnología Informática y Computación, Universidad de Alicante, 2013.
- [18] Jorge del Valle Arias, Rodrigo Palacios Marugán, «Trabajar la tierra con redes de sensores inalámbricos,» BIT, del Colegio Oficial de Ingenieros de Telecomunicación de España (COIT), vol. 1, nº 188, pp. 67-70, 2012.
- [19] J. Quiza. , T. Castelblanco. , D. González, «Sistema Web de Identificación de Ganado Bovino y Monitoreo en Tiempo Real de su Temperatura y Frecuencia Cardiaca Usando Redes de Sensores Inalámbricos,» *ENGI Revista Electrónica De La Facultad De Ingeniería* , vol. 1, nº 1, 2012.
- [20] C. Perkins, Elizabeth M. Belding-Royer, S. Das, «Ad hoc On-Demand Distance Vector (AODV) Routing,» 2003.
- [21] T. Shah, N. Javaid, T. N. Qureshi, «Energy Efficient Sleep Awake Aware (EESAA),» COMSATS Institute of Information Technology, 44000, Islamabad, Pakistan, 2012.
- [22] Rahul C. Shah, Jan M. Rabaey, «Energy Aware Routing for Low Energy Ad Hoc Sensor Networks,» *Wireless Communications and Networking Conference*, 2002.
- [23] S.Ganesh, R.Amutha, «Efficient and Secure Routing Protocol for Wireless Sensor Networks,» Sathyabama University, SSN College of Engineering, Chennai, Tamil Nadu, India, 2013.
- [24] V. Karthikeyan, A.Vinod, P. Jeyakumar, «An Energy Efficient Neighbour Node Discovery Method for Wireless Sensor Networks,» 2014.
- [25] Mostafa SAADI, Moulay Lahcen HASNAOUI, Abderrahim BENI HSSANE, Said BENKIRANE, Mohamed LAGHDIR, «Energy-Aware Scheme used in Multi-level Heterogeneous Wireless Sensor Networks,» 2013.
- [26] M. Tahir, N. Javaid, A. Iqbal, Z. A. Khan, and

- N. Alrajeh, «On Adaptive Energy-Efficient Transmission in WSNs,» International Journal of Distributed Sensor Networks, 2013.
- [27] Amit Sharma, Kshitij Shinghal, Neelam Srivastava, Raghuvir Singh, «Energy Management for Wireless Sensor Network,» de International Journal of Advances in Engineering & Technology., 2011.
- [28] Q. Nadeem, M. B. Rasheed, N. Javaid, Z. A. Khan, Y. Maqsood, A. Din, «M-GEAR: Gateway-Based Energy-Aware Multi-Hop Routing Protocol for WSNs,» COMSATS Institute of Information Technology, Faculty of Engineering, Dalhaousie University, Islamabad, Pakistan; Halifax, Canada, 2013.
- [29] Y. Khan, N. Javaid, M. J. Khan, Y. Ahmad, M. H. Zubair, S. A. Shah, «LPCH and UDLPCH: Location-aware Routing Techniques in WSNs,» Dept of Electrical Engineering, COMSATS Institute of IT; University of Oulu, Islamabad, Pakistan; Oulu, Finland, 2013.
- [30] Eduardo O. Sosa, Diego Alberto Godoy, Rebeca Díaz Redondo, Rodolfo M. Lilli, Juan de Dios Benitez, «Localización Geográfica de Ganado Utilizando Modelos de Propagación de Señal y XBee,» de XVI Workshop de Investigadores en Ciencias de la Computación, Ushuaia, Tierra del Fuego, 2014.
- [31] Raúl Luft, Darios Sosa, Diego A. Godoy, Eduardo O. Sosa, Juan D. Benitez, «Validando aplicaciones para Ciudades Inteligentes - Recolección de Residuos Urbanos».
- [32] C. Ltd, «Ubuntu,» [En línea]. Available: <http://www.ubuntu.com/download/desktop>. [Último acceso: 04 10 2015].
- [33] O. Ltd., «OMNeT++,» [En línea]. Available: <https://omnetpp.org/>. [Último acceso: 04 10 2015].
- [34] A. A. Quintana, «GitHub,» [En línea]. Available: <https://github.com/aarizaq/inetmanet-2.0>. [Último acceso: 04 10 2015].
- [35] OMNeT++ developers, [En línea]. Available: <https://inet.omnetpp.org/>. [Último acceso: 04 10 2015].
- [36] Uppsala University, [En línea]. Available: <http://moment.cs.ucsb.edu/AODV/>. [Último acceso: 04 10 2015].
- [37] Uppsala University, [En línea]. Available: <http://www.uu.se/>. [Último acceso: 04 10 2015].
- [38] The MathWorks, Inc., [En línea]. Available: <http://www.mathworks.com/products/matlab/>. [Último acceso: 04 10 2015].
- [39] F. Ren, J. Zhang, T. He, C. Lin y S. Ren, «EBRP: Energy-Balanced Routing Protocol for Data Gathering in Wireless Sensor Networks,» IEEE Transactions on Parallel and Distributed Systems, vol. 22, nº 12, pp. 2108-2125, 2011.

Clasificación por Enterotipos y Grupos Ortólogos del Microbioma Humano con Métodos No Supervisados

Cristóbal R. Santa María*, Victoria Santa María**, Laura Ávila*,
Juan Otaegui*, Marcelo Soria***

*DIIT-UNLaM, **Instituto Lanari-FMed-UBA, ***FAUBA

Florencio Varela 1903 San Justo Pcia. de Buenos Aires 54-011-44808952

csantamaria@unlam.edu.ar vcetrntmr@hotmail.com

laura_avila75@yahoo.com.ar soria@agro.uba.ar

juancarlosotaegui@yahoo.com.ar

Resumen

Se relatan las tareas llevadas a cabo por el Grupo de Investigación y Desarrollo en Data Mining del Departamento de Ingeniería e Investigaciones Tecnológicas de la UNLAM durante el año 2015 en el marco del Proyecto de Incentivos C169 “Aplicaciones de Data Mining al Estudio del Microbioma Humano”. Se detallan las pruebas realizadas con el software SUPERFOCUS y la base de datos genéticos SEED, para desarrollar los análisis taxonómicos y funcionales que prepararan la información de las secuencias microbiómicas para procesarla con algoritmos de data mining. Se explicitan los aspectos teóricos y prácticos de la aplicación de estos algoritmos sobre conjuntos de prueba. Se analiza la interpretación clínica dada a los resultados y finalmente se describen los cursos de acción para continuar con la investigación durante 2016.

Palabras Claves: ADN, Microbioma, Enterotipos, Grupo Ortólogo, Clasificación.

Contexto

En esta línea de investigación se intenta construir un algoritmo de clasificación de estadios de desarrollo de cáncer de colon y enfermedad de Crohn basado en la información aportada por el ADN constituyente del microbioma humano, en particular el intestinal. A partir de la secuenciación del ADN microbiano presente en el intestino, cada secuencia

genética es una instancia en una base de datos sobre la que es posible aplicar procedimientos de aprendizaje no supervisado para agrupar las secuencias correspondientes a un gen marcador por especies u otros taxones más generales. Sin embargo tales categorizaciones suelen tener un sesgo, respecto de la caracterización clínica, que es producto de la secuenciación misma y de las técnicas que se aplican previas al agrupamiento. Por tal motivo conviene explorar agrupamientos de todas las secuencias genéticas, y no ya solo las de un gen marcador, realizados también en forma no supervisada, de acuerdo a la función que les corresponda en el metabolismo y que resulta distinta en la salud y en cada estadio de la enfermedad. Se espera que el estudio realizado interrelacionando ambos tipos de agrupamientos proporcione categorías de clasificación estables y compatibles con las caracterizaciones clínicas de los estadios de salud y enfermedad. Con tales categorías se intentará a continuación aplicar métodos de aprendizaje supervisado como ensambles de árboles de decisión para obtener un clasificador que colabore en la clínica de prevención, diagnóstico o pronóstico. Se pretende además elaborar como resultado de los pasos descriptos una "pipeline" para investigar la aplicación de técnicas de data mining al microbioma humano en el caso de una enfermedad en general.

Introducción

Con vistas a la línea de investigación que el presente proyecto abre y que incluye la posibilidad futura de usar muestras propias, se decidió adaptar un servidor con el sistema BioLinux que posee una fácil interacción con paquetes de software libre utilizados en la investigación en biología computacional. Ocasionalmente se usaron también máquinas virtuales armadas en la nube para llevar adelante los procesos.

Al tener presente el objetivo de desarrollar una “pipeline” que en primer término instrumente una clasificación de microbiomas ajustada según criterios clínicos y luego utilice las categorías obtenidas para predecir en forma de diagnóstico o pronóstico las enfermedades estudiadas, se decidió llevar a cabo un ensayo sobre los procedimientos más citados para estas acciones. Para cobrar idea del volumen de datos a utilizar se comenzó estableciendo que espacio promedio, calculado en bytes, ocupa una secuencia. Los relevamientos de ADN total de microbiomas producen fragmentos cortos de ADN. Dependiendo de la tecnología de secuenciación y la configuración de los equipos se puede controlar el tamaño y cantidad de estos fragmentos. Con los equipos MiSeq y HiSeq de Illumina, los más usados hoy en día, las secuencias de ADN pueden tener entre 150 y 300 bases de largo y la cantidad puede llegar a varios millones de secuencias por corrida. Es común en el análisis de microbiomas tener unos 30 millones de secuencias por cada muestra analizada. Esto determina que sean críticos los tiempos de ejecución de los programas utilizados para el análisis.

Existen diferentes tipos de análisis para estudios de secuenciación metagenómica, por ejemplo, obtener “contigs”, que son ensamblajes de secuencias en un fragmento bastante mayor y que permite la reconstrucción de genes completos, genomas parciales o, incluso, genomas

completos. Otras aproximaciones posibles son la asignación taxonómica de las secuencias, es decir, determinar con la mayor precisión posible a que especie, género, familia, etc. de microorganismos pertenecen las secuencias obtenidas; o determinar una asignación funcional, que consiste en encontrar para las secuencias que codifican proteínas, de qué tipo son y en qué actividades celulares participan.

Líneas de Investigación, Desarrollo e Innovación

La línea de trabajo elegida se enfoca en las asignaciones taxonómicas y funcionales [1]. Para esto se realizaron pruebas con un software de reciente desarrollo, SUPERFOCUS, que efectúa la determinación taxonómica y la asignación funcional. Para esta última actividad el software utiliza la base de datos SEED [2] que asigna una función a cada proteína. SUPERFOCUS corre en plataformas Linux y está diseñado para aprovechar las características multiprocesador de las máquinas modernas. Las pruebas iniciales se realizaron en el servidor BioLinux con ocho procesadores Intel-I7 y 16 GB de memoria RAM. Se continuaron las pruebas en el servicio de computación en la nube con una máquina virtual Linux. Se determinó que con archivos de 500,000 secuencias el proceso se completaba con éxito en todas las pruebas. Se diseñó entonces un script de comandos Unix que particiona el archivo a procesar en bloques de 500,000 secuencias cada uno y las envía a SUPERFOCUS. Se procesaron cinco muestras diferentes, cada una con unos 26 millones de secuencias.

Las salidas de SUPERFOCUS son varios archivos de texto, uno con las asignaciones funcionales, tres con la información de subsistemas y otro más con la información taxonómica.

El procedimiento experimentado permite entonces conocer la distribución de frecuencias de genes o funciones de cada

paciente con la cual se realizará la clasificación en clusters de sus respectivos microbiomas.

Resultados y Objetivos

En [3] se analiza la influencia de la dieta en el cáncer de colon mediada por la composición de la microbiota a partir de muestras de población afroamericana (AA) y africana nativa (NA). Se miden las diferencias de microbiota de los AA y los NA. La representación numérica de cada microbioma intestinal se realiza entonces por un vector donde cada componente expresa la abundancia de cada microorganismo en él.

En el caso de las muestras AA y NA se trata de dos matrices integradas por las filas que representan los microbiomas de los individuos integrantes. El número de componentes del vector-microbioma es de unos cientos. Para visualizar en el caso presentado las diferencias entre los microbiomas AA y NA se apela a una reducción de variables por componentes principales resultando la Figura 1. De acuerdo a lo expuesto hasta aquí el trabajo computacional consiste en obtener los datos secuenciados de una muestra integrada por varios microbiomas

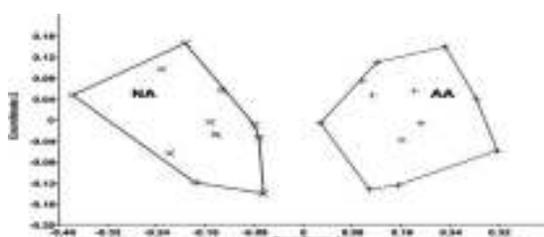


Figura 1

Cada microbioma de esta muestra debe cotejarse con una base de datos correspondiente a un gen marcador para encontrar la distribución de frecuencias de los microorganismos identificados por tal gen. Alternativamente el conjunto de secuencias del microbioma puede compararse con otra base de datos de

funciones genéticas para agrupar los genes integrantes por función y así obtener la distribución de frecuencias según las funciones metabólicas que las secuencias integrantes revelan. [1]. En cualquier caso los microbiomas individuales pueden agruparse en clusters. Los conjuntos obtenidos clasifican a los individuos según características clínicas cuyo valor debe ser sopesado desde el punto de vista médico. Si la clasificación tiene significancia clínica puede utilizarse con métodos predictivos tales como árboles de decisión para establecer diagnósticos o pronósticos en pacientes aún no clasificados [4]

A partir de [3] y a efecto de estudiar la metodología empleada, se obtuvieron dos conjuntos de datos A y B que agrupan a los microbiomas intestinales de 33 pacientes. Cada fila del conjunto A representa el microbioma de un paciente obtenido en base al gen marcador. Para el conjunto B cada fila representa el microbioma de un paciente y cada columna una característica o grupo funcional hallada.

Se continuó realizando la clasificación en clusters. Para medir la distancia entre dos muestras, dos microbiomas cuyas distribuciones de especies o géneros son conocidas se consideró que cada paciente está representado por una distribución de frecuencias estadísticas. Hay que medir si las dos distribuciones se parecen a efecto de terminar ubicándolas en clusters. Para ello se utilizó la distancia de Jensen-Shanon [5]

$$D_{PQ} = \left(\sum_{i=1}^N p_i \log p_i + q_i \log q_i \right)^{\frac{1}{2}}$$

Esta distancia es la que usa el algoritmo PAM (Partitioning around medoids) utilizado para este caso directamente de la biblioteca cluster de R. El medoide es el elemento para el cual la disimilitud promedio con todos los objetos en el conglomerado es mínima. Se estudió también, con la idea de programarlo como alternativa más rápida el algoritmo

propuesto en [6] que trabaja con medoides pero en forma similar a k-means.

Para definir el número óptimo de clusters se calculó el índice de Calinski-Harabasz :

$$CH = \frac{B_k}{\frac{W_k}{n-k}}$$

Aquí B_k es la suma de las distancias al cuadrado de todos los elementos i y j que no pertenecen al mismo cluster, W_k es la suma de los cuadrados de las distancias de todos los elementos i y j que pertenecen al mismo cluster, n es el número de elementos a clasificar y k la cantidad seleccionada de clusters. Utilizando el comando `nclusters` de la biblioteca `clusterSim` de R se pudo programar el testeo de CH para distintos valores de k a fin de hallar la cantidad óptima de clusters. En la Figura 2 se ve un ejemplo para el que se ha utilizado el conjunto A. El número de clusters $k = 3$ produce el mayor valor del índice CH.



Figura 2

Con el número de clusters obtenido se corre el algoritmo PAM para obtener los enterotipos en cuestión. Desde el punto de vista computacional la consistencia de tal agrupamiento se mide con el índice Silhouette [7] Sin embargo luego de las pruebas de agrupamientos realizadas se requirió evaluar el aspecto clínico de los resultados obtenidos.

Hay más de 1000 especies de microbios que viven en el intestino humano y conforman el microbioma. Este juega un rol importante en la protección del huésped contra patógenos, modula la inmunidad, regula procesos metabólicos y es incluso considerado en algunos casos como un “órgano endocrino”. Secuenciando 16s rADN ha sido sugerido

que el microbioma intestinal puede dividirse en tres diferentes “enterotipos” [3], cada uno de ellos pueden ser identificados por la variación en los niveles de alguno de los siguientes tres géneros: Bacteroides (enterotipo 1) Prevotella (enterotipo 2) y Rominococcus (enterotipo 3). Sin embargo estos enterotipos no se definen tan claramente como por ejemplo los grupos sanguíneos, no parecen ser distintos con respecto a su riqueza funcional y no han podido ser correlacionados a características poblacionales tales como la edad, el IMC o la nacionalidad.

El estudio de biomarcadores funcionales arrojó mejores resultados, hallándose correlación entre algunos de ellos y propiedades del huésped lo que puede ser útil para el diagnóstico y prevención de enfermedades. El resultado final de este análisis se utiliza para describir un número potencial de funciones y su relativa abundancia en el metagenoma.

Para realizar estos estudios en vez de dividir al microbioma en diferentes enterotipos por clasificación taxonómica se lo divide en grupos ortólogos (OG) que codifican para distintas enzimas o proteínas pertenecientes a distintas vías metabólicas. [3] Para la Enfermedad de Crohn se ha encontrado que su manifestación está asociada a una disminución de las especies bacterianas (Bacteroides y Clostridium) en la mucosa colónica y aumento de otras tales como las fusobacterias. Sin embargo esto no se refleja en las muestras tomadas de la materia fecal [8]. La etiología del cáncer colorectal es multifactorial. Se ha encontrado una abundancia relativa de Bacteroidaceae, Streptococcaceae, Fusobacteriaceae, Peptostreptococcaceae, Veillonellaceae y Pasteurellaceae y una disminución en los niveles de Lachnospiraceae, Ruminococcaceae y Lactobacillaceae en los pacientes con

cáncer colorrectal [9]. Además el microbioma induce un estado de inflamación crónica y genera metabolitos reactivos y carcinógenos que contribuyen al desarrollo del cáncer colorrectal. Se ha trabajado sobre la hipótesis de que la utilización de biomarcadores del microbioma y el análisis de otros factores de riesgo pueden contribuir al screening de cáncer colorrectal. En este caso se encuentran diferencias significativas entre los sujetos normales y los distintos estadios de cáncer colorrectal [10]. También estas disbiosis favorecen la prevalencia de bacterias dentro del tumor que poseen genes altamente virulentos que contribuyen al desarrollo de la enfermedad [9]. La división según características taxonómicas no parece ser la más adecuada dado que no se refiere a la funcionalidad de las bacterias y tiene una amplia variación inter-individuos, por lo que el clustering debe realizarse según grupos de co-abundancia que combinen a los distintos tipos de bacterias en función de las vías metabólicas asociadas a cada patología. Ya se ha analizado que existen vías metabólicas que resultan protectoras, otras que son desencadenantes de la enfermedad y otras que la perpetúan, probablemente esta sea la forma de empezar la clasificación. Por otro lado la flora de la materia fecal no es igual a la flora de la mucosa colónica por lo que el estudio sería más representativo si se tomaran muestras a partir de biopsia de mucosa colónica.

Formación de Recursos Humanos

En el grupo participan un Magister y un Especialista en Explotación de Datos, un Doctor en Biología, dos Médicos, dos Ingeniero en Sistemas, un Matemático y un alumno de la carrera de Informática. Actualmente hay en desarrollo una tesis de maestría.

Bibliografía

- [1] Ngom-Bru, Catherine and Barretto, Caroline. Gut microbiota: methodological aspects to describe taxonomy and functionality. Briefings in Informatics. Vol3 NO 6. 747-750
- [2] <http://theseed.org>
- [3] Arumugam, M et al. Enterotypes of the human gut microbiome. Nature 2011 may 12; 473(7346): 174-180. doi:10.1038/nature09944
- [4] Junhai et al. Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. Am. J. Clin. Nutr. 2013. 98. 11-120.
- [5] Endres, D y Schindeling, J. A New Metric for Probability Distributions. IEEE. Transactions on Information Theory. Vol. 49 NO.7. 2003.
- [6] Hae-Sang Park, Chi-Hyuck Jun. A simple and fast algorithm for K-medoids clustering. Expert Systems with Applications 36 (2009) 3336-3341.
- [7] Rousseeuw, P. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20 (1987) 53-65.
- [8] Ray K. IBD. Understanding gut microbiota in new-onset Crohn's disease. Nat Rev Gastroenterol Hepatol [Internet]. Nature Publishing Group; 2014;11(5):268.
- [9] Burns MB, Lynch J, Starr TK, Knights D, Blekhman R. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. Genome Med [Internet]. 2015;7(1):55.
- [10] Zackular JP, Rogers MAM, Ruffin MT, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. Cancer Prev Res (Phila) [Internet]. 2014;7(11):1112-21.

Extracción de Conocimiento en el Cursado del Ciclo Común de Articulación de Carreras de Ingeniería

Oscar Eduardo Quinteros¹, Ana Funes², Hernán César Ahumada¹

¹Universidad Nacional de Catamarca

Maximio Victoria 55, 4700 Catamarca, Argentina

{oequinteros,hcahumada}@tecno.unca.edu.ar

²Universidad Nacional de San Luis

Ejército de los Andes 950, 5700 San Luis, Argentina

afunes@unsl.edu.ar

Resumen

Sobre datos de regularidades y exámenes rendidos de las cátedras pertenecientes al Ciclo Común de Articulación (CCA) de las carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas, se puede aplicar métodos de minería de secuencias para obtener el recorrido académico de los alumnos y de esta manera poder descubrir reglas de secuencias (patrones) frecuentes en la historia académica de cada uno de ellos. El análisis de las secuencias obtenidas permitirá caracterizar a los grupos de alumnos con diferentes rendimientos académicos y grado de avance en los estudios. La interpretación de los resultados posibilitará conocer las materias que mayor tiempo insumen hasta que son aprobadas. El conocimiento obtenido brindará la posibilidad de entender mejor un aspecto de la problemática del abandono temprano de los estudios universitarios.

Palabras clave: Minería de Datos, Minería de Secuencias, Rendimiento Académico, Deserción.

Contexto

Esta línea de trabajo se viene llevando a cabo como un trabajo colaborativo entre proyectos de investigación de dos universidades nacionales: el Proyecto de Incentivos código 22/F222 *Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software*, de la Universidad Nacional de San Luis, y el proyecto acreditado por SECyT - UNCA. 02/J715 - *Hacia La Búsqueda De Un Sistema De Alerta Temprana De Deserción En Carreras De Ingeniería*, que se desarrolla en el Departamento de Formación Básica de la Facultad de Tecnología y Ciencias Aplicadas (Universidad Nacional de Catamarca).

En particular, esta propuesta de investigación se desarrolla como una tesis de posgrado de la carrera Maestría en “Ingeniería de Software” de la Universidad Nacional de San Luis.

Introducción

Las autoridades y equipos docentes de universidades se plantean la mejora continua de la calidad de los procesos y servicios que ofrecen, de esta manera buscan incrementar su contribución al desarrollo social y económico de su entorno. Un estudio realizado por

Ahumada [3], indica que las carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca (U.N.Ca.) tienen estadísticas de deserción y desgranamiento de magnitud significativa y se producen, fundamentalmente, en los primeros años de estudios, en donde la mayoría de los análisis descriptivos de este fenómeno revelan demasiado tarde el abandono de estudiantes en diferentes tramos de la carrera.

La Facultad de Tecnologías y Ciencias Aplicadas¹, de la Universidad de Catamarca, junto a otras facultades del país, se ha adherido a un convenio con el fin de dictar un conjunto común de materias, denominado Ciclo Común de Articulación (CCA) [14], en la etapa inicial de las carreras de ingeniería. El estudio del rendimiento académico de los alumnos en el cursado del CCA puede brindar información útil para disminuir la deserción temprana de los alumnos en las carreras.

El sistema de gestión de información académica de las Universidades Nacionales denominado, SIU-GUARANI, lleva un registro actualizado de los exámenes rendidos por los alumnos, junto con otros datos académicos referidos básicamente a inscripciones de materias para cursar y actas de regularidades. En los últimos años, la Facultad de Tecnología y Ciencias Aplicadas, ha implementado sistemas de gestión que integran procesos y áreas, donde estos sistemas producen datos que se almacenan en bases de datos, con dimensiones considerablemente grandes y diversidad de temas. Los trabajos realizados por Ahumada [3] y Pautsch [11] muestran la aplicación del uso de Minería de Datos (MD) en estos

entornos educativos, particularmente enfocados en la identificación de las causales de abandono.

Más específicamente, existe una rama en la MD que se denomina Minería de Datos Educativos (EDM), la cual es una disciplina relacionada con el desarrollo de métodos para extraer información útil a partir de los datos que se generan en los entornos educativos, y utilizarla para mejorar dicho entorno. Romero [15] expresa que la información así obtenida se convierte en el insumo indispensable para la toma de decisiones. La EDM tiene varios objetivos, tales como los descriptos por Jindal [9]. Uno de estos objetivos es el estudio, desde el punto de vista académico, de la performance del alumno. La aplicación de MD para extracción de conocimiento desde los datos que gestionan las Universidades es factible, ya que desde hace varios años, las Universidades Nacionales vienen trabajando en la implementación de sistemas de gestión o transaccionales que integran procesos y áreas. Estos sistemas producen datos que se almacenan en bases de datos, con dimensiones considerablemente grandes y diversidad de temas, como indica Menéndez en [10].

Los métodos de MD principalmente usados en el EDM se dividen en 2 grupos, los orientados a la verificación y los orientados al descubrimiento: dentro de estos últimos tenemos a su vez métodos de Clasificación, Estadísticas, Agrupamiento, Predicción, Redes Neuronales, Minería de Reglas de Asociación y Minería Web [9].

Galindo[6] también describe categorías de métodos de EDM, donde una de ellas es la Minería de Relaciones y, dentro de ésta, los métodos de minería de patrones de secuencias. La Minería de Secuencia [1] consiste en la búsqueda de patrones secuenciales frecuentes en una base de datos de eventos con fecha y hora [13][5].

¹<http://www.tecno.unca.edu.ar>

Existen además, diversos métodos para minar secuencias, Ahola [2] explicita varios, tales como SPADE [16], o PrefixSpan [12], entre otros. Como antecedente en la aplicación de minar secuencias sobre datos provenientes de entornos educativos tenemos el trabajo de Guerra [7] en donde usa los métodos SPAM [4] y PexSPAM [8] para descubrir patrones comunes en secuencias de ejercicios parametrizados usados en las herramientas de aprendizaje on-line.

Ante lo expuesto, se considera que es posible obtener los datos relacionados con exámenes rendidos y modelarlos como una serie y luego aplicar métodos para minar las secuencias. El sistema SIU-GUARANI lleva un registro de los exámenes rendidos por los alumnos durante el cursado del CCA, que combinado con otros datos académicos puede conformar un dataset para ser analizado por métodos de MD.

Líneas de Investigación, Desarrollo e Innovación

La línea de investigación aquí presentada trata del un tema surgido para el desarrollo de una tesis de Maestría en Ingeniería de Software de la Universidad Nacional de San Luis.

El proyecto comprende la investigación de Minería de Datos en base a registros académicos obtenidos del SIU-GUARANI del cursado de las cátedras que se dictan en el CCA de las carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas de la U.N.Ca. Se busca, mediante la aplicación de métodos de minería de datos, caracterizar y analizar el recorrido académico de los estudiantes, con la finalidad de detectar patrones que describan la demora en la terminación del

cursado, que puedan derivar posteriormente en abandono de los estudios universitarios.

Resultados y Objetivos

Objetivo General

Caracterizar y analizar el recorrido académico de exámenes rendidos por alumnos del Ciclo Común de Articulación (CCA) de las Carreras de la Facultad de Tecnologías y Ciencias Aplicadas de la Universidad Nacional de Catamarca (U.N.Ca.).

Objetivos Específicos

- Descubrir reglas que caractericen la demora en la terminación del cursado del CCA.
- Encontrar características en una secuencia de exámenes rendidos que permitan explicar algunas de las causales que puedan generar que un alumno tenga una demora excesiva.
- Descubrir relaciones entre las materias y los patrones de demoras en la rendición de los exámenes.
- Buscar patrones frecuentes en el conjunto de exámenes rendidos de alumnos con mejor y peor rendimiento.
- Brindar elementos para poder realizar un análisis institucional y que permita tomar decisiones para definir nuevas estrategias dentro de la institución, como por ejemplo re-definir planes de estudio, o realizar cursos de nivelación.

Formación de Recursos Humanos

El presente trabajo es la base para el desarrollo de una tesis de posgrado de la

carrera “Maestría en Ingeniería de Software” de la Universidad Nacional de San Luis. La misma está siendo llevada adelante por el Licenciado en Sistemas de Información Oscar Eduardo Quinteros, y siendo dirigida por la investigadora de la Universidad Nacional de San Luis Magister Ana Funes, y codirigido por el investigador de la Universidad Nacional de Catamarca, Dr. Hernán Cesar Ahumada. El trabajo de tesis se encuentra en la etapa de elaboración del proyecto de tesis.

Referencias

[1] Rakesh Agrawal and Ramakrishnan Srikant. Mining sequential patterns. In Data Engineering, 1995. Proceedings of the Eleventh International Conference on, pages 3_14. IEEE, 1995.

[2] Jussi Ahola. Mining sequential patterns. 2001.

[3] Hernán César Ahumada, Hugo Rubén Dip, Carlos Gabriel Herrera, and Juan Carlos Leguizamón Almendra. Minería de datos para un sistema de alerta temprana de deserción en carreras de ingeniería. In Anales XVII Workshop de Investigadores en Ciencias de la Computación WICC 2015), Abril 16-17, 2015, Salta, Argentina. Red UNCI, 2015.

[4] Jay Ayres, Jason Flannick, Johannes Gehrke, and Tomi Yiu. Sequential pattern mining using a bitmap representation. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 429_435. ACM, 2002.

[5] Dante Conti and Fco Javier Martínez de Pisón Ascacíbar. Reglas de asociación en series temporales: panorama referencial y tendencias.

[6] Álvaro Jiménez Galindo and H García. Minería de datos en la educación. Universidad Carlos III, 2010.

[7] Julio Guerra, Shaghayegh Sahebi, Yu-Ru Lin, and Peter Brusilovsky. The problem solving genome: Analyzing sequential patterns of student work with parameterized exercises. In Educational Data Mining 2014, 2014.

[8] Joshua Ho, Lior Lukov, and Sanjay Chawla. Sequential pattern mining with constraints on large protein databases. In Proceedings of the 12th International Conference on Management of Data (COMAD), pages 89_100, 2005.

[9] Rajni Jindal and Malaya Dutta Borah. A survey on educational data mining and reserarch trends. International Journal of Database Management Systems, 5(3), 2013.

[10] Mariano Menéndez and María de Lujan Gurmendi. Sistemas para la toma de decisiones en el ámbito universitario. In 6o Simposio Argentino De Informática En El Estado, 2012.

[11] JGA Pautsch. Minería de datos aplicada al análisis de la deserción en la carrera de analista en sistemas de computación. Misiones, Argentina: Universidad Nacional de Misiones, Facultad de Ciencias Exactas, Químicas y Naturales, 2009.

[12] Jian Pei, Jiawei Han, Behzad Mortazavi-Asl, Jianyong Wang, Helen Pinto, Qiming Chen, Umeshwar Dayal, and Mei-Chun Hsu. Mining sequential patterns by pattern-growth: The pre_xspan approach. Knowledge and Data Engineering, IEEE Transactions on, 16(11):1424_1440, 2004.

[13] Prof Pinkal Shah and AK Dua. Algorithm for sequence mining using gap constraints. International Journal of Engineering Research and Development, pages 37_49, 2014.

[14] Facultad de Tecnologías y Ciencias Aplicadas. Historia - Facultad - CCA. [Web:www.tecno.unca.edu.ar/index.php/facu/historia; accedido el 29-02-2016].

[15] Cristobal Romero, Sebastian Ventura, Mykola Pechenizkiy, and Ryan SJD Baker. Handbook of educational data mining. CRC Press, 2010.

[16] Mohammed J. Zaki. Spade: An efficient algorithm for mining frequent sequences. Mach. Learn., 42(1-2):31_60, January 2001.

Grandes Datos y Algoritmos Eficientes para Búsquedas de Escala Web

Gabriel H. Tolosa^{1,2}, Santiago Bancho¹, Esteban A. Ríssola¹,
Tomás Delvechio¹ y Esteban Feuerstein²
{tolosoft, sbancho, earissola, tdelvechio}@unlu.edu.ar; efeurest@dc.uba.ar

¹Departamento de Ciencias Básicas, Universidad Nacional de Luján

²Departamento de Computación, FCEyN, Universidad de Buenos Aires

Resumen

El acceso a la información en tiempo y forma es un factor esencial en muchos procesos que ocurren en dominios diferentes: la academia, la industria, el entretenimiento, entre otros. En la actualidad, el enfoque más general para acceder a la información en la web es el uso de motores de búsqueda de gran escala. Éstos sistemas enfrentan constantes desafíos debido al crecimiento explosivo de contenido en la web y también de la cantidad de nuevos usuarios. Principalmente, aparecen nuevas necesidades de almacenamiento y procesamiento para satisfacer estrictas restricciones de tiempo: las consultas deben ser respondidas en pequeñas fracciones de tiempo, típicamente, milisegundos.

Esta problemática tiene aún muchas preguntas abiertas y – mientras se intentan resolver cuestiones – aparecen nuevos desafíos. Existen necesidades puntuales de los servicios que recolectan y utilizan esta información tal como nuevas estructuras de datos y algoritmos altamente eficientes lo que brinda oportunidades únicas para avances científico/tecnológicos en áreas como algoritmos, estructuras de datos, sistemas distribuidos y procesamiento de datos a gran escala, entre otras.

En este proyecto se estudian, proponen, diseñan y evalúan estructuras de datos y algoritmos eficientes junto con el análisis de grandes datos que permitan aumentar procesos internos de un motor de búsqueda con el objetivo de mejorar su performance y escalabilidad.

Palabras clave: motores de búsqueda, estructuras de datos, algoritmos eficientes, grandes datos.

Contexto

Esta presentación se encuentra enmarcada en el proyecto de investigación “Algoritmos Eficientes y Minería Web para Recuperación de Información a Gran Escala” del Departamento de Ciencias Básicas (UNLu) en el cual los autores son integrantes (Disp. CD-CB N° 327/14). Complementariamente, el primer autor desarrolla su tesis de doctorado en el Depto. de Computación de la FCEyN (UBA) en esta temática.

Introducción

En los últimos años el número y complejidad de documentos en la web ha crecido exponencialmente, convirtiéndola en el mayor repositorio de información en el mundo. Esto crea nuevas necesidades de almacenamiento, procesamiento y búsquedas, expandiendo los límites del trabajo en una sola máquina y unos pocos algoritmos al trabado distribuido, paralelo y altamente eficiente. En este escenario existen por un lado, necesidades puntuales de los servicios que recolectan y utilizan información de la más diversa y compleja y por el otro, aparecen oportunidades únicas para avances científico/tecnológicos en áreas como algoritmos, estructuras de datos, sistemas distribuidos y procesamiento de datos a gran escala.

El acceso a la información en tiempo y forma es un factor esencial en muchos procesos que ocurren en dominios diferentes: la academia, la industria, el entretenimiento, entre otros. En la actualidad, el enfoque más general para acceder a la información en la web es el uso de motores de búsqueda, a partir de consultas basadas en las necesidades de información de los usuarios. De forma simple, los motores

de búsqueda intentan satisfacer la consulta de los usuarios realizando procesos de recuperación sobre una porción del espacio web que “conocen”, es decir, que han recorrido, recopilado y procesado [2]. Esto es así dado que existen porciones de la web a las cuales no se puede acceder debido a su dinámica o restricciones de acceso. De aquí que su tamaño no se puede determinar de manera precisa¹. Además, el número de usuarios crece permanentemente [25] y éstos no solamente *buscan* en la web para satisfacer sus necesidades de información sino que - además - realizan tareas cotidianas (por ejemplo, organizar un viaje, comprar cosas, etc.). Además, todas estas aplicaciones operan con estrictas restricciones de tiempo: las consultas deben ser respondidas en pequeñas fracciones de tiempo, típicamente, milisegundos. Los motores de búsqueda se han convertido en herramientas indispensables en la Internet actual y las cuestiones relacionadas con su eficiencia (escalabilidad) y eficacia son temas de muy activa investigación [4].

En su arquitectura interna, las máquinas de búsqueda de gran escala presentan un grado de complejidad desafiante [6], con múltiples oportunidades de optimización. Como la web es un sistema dinámico que en algunos casos opera en tiempo real, las soluciones existentes dejan de ser eficientes y aparecen nuevas necesidades. Paralelamente, en los últimos años se ha popularizado el uso de técnicas estadísticas y de *machine learning* para lograr extraer modelos útiles a partir de repositorios de datos [14] complejos. Esta disciplina, conocida como minería de datos (o minería web en este contexto), es una etapa de un proceso más complejo, el de descubrimiento de conocimiento, que puede ser altamente útil en el ámbito de los motores de búsqueda [24].

Además, el crecimiento de los repositorios y de las diferentes fuentes de generación de información (redes sociales, sensores, etc.) han agregado mayor complejidad y la necesidad de dar respuestas en tiempo real. Se ha redoblado la apuesta y gran parte de los problemas que se trataban desde la óptica de la minería de datos pasaron a ser problemas de “Grandes Datos” [27] (Big Data), donde las soluciones a éstos son significativamente más complejas ya que los volúmenes de información son muy grandes,

¹De acuerdo al sitio World Wide Web Size (<http://www.worldwidewebsize.com/>), se estiman unos 48.000 millones de documentos en la web superficial (accedida por los motores de búsqueda)

llegan de manera continua y requieren respuestas en tiempo real [20].

Los problemas de Grandes Datos requieren de soluciones complejas que sobre arquitecturas que puedan escalar de manera flexible [21], tanto en cómputo como almacenamiento. En las grandes organizaciones el conocimiento y manejo de grandes volúmenes de datos permite tomar mejores decisiones a partir de evidencia, es decir, algunas soluciones surgen de los datos (*data driven*) y no de la intuición [16].

Las técnicas para descubrimiento de conocimiento son transversales a cualquier disciplina científica, por lo que se considera que existe un amplio abanico de soluciones de optimización aún no exploradas para los motores de búsqueda a gran escala que pueden ser tratadas siguiendo una metodología de minería de datos. Principalmente, en problemáticas que abarcan desde el análisis profundo de query logs en buscadores y query recommendation hasta políticas para la optimización de caches [24].

Líneas de investigación y desarrollo

En este proyecto se continúan líneas de I+D del grupo que incorporan análisis de grandes datos para rediseñar algunos procesos internos de un motor de búsqueda web que permitan aumentar sus prestaciones. Existen oportunidades de investigación en temas poco explorados por la comunidad científica que permiten mejorar y/o rediseñar los algoritmos internos y las estructuras de datos usadas principalmente para recuperación de información de gran escala. En especial, las líneas de I+D principales son:

a. Estructuras de Datos

1. Distribuidas: Los sistemas de búsqueda en texto utilizan como estructura de datos básica un índice invertido. De forma simple, este índice está formado por un vocabulario (V) con todos los posibles términos de búsqueda y un conjunto de *posting lists*, L , con información acerca de los documentos donde aparece cada término junto con datos extra (por ejemplo, la frecuencia del término i en el documento j). Como los sistemas de búsqueda a gran escala se ejecutan en clusters de computadoras, es necesario distribuir los documentos entre los nodos. Para ello, los dos enfoques clásicos [2] son:

* **Particionado por documentos:** La colección, C , es dividida entre P procesadores, los cuales almacenan solo una porción del índice $\frac{C}{P}$.

* **Particionado por términos:** Cada nodo mantiene información de las listas de posting completas de solamente un subconjunto de los términos. El vocabulario V es dividido entre los P nodos y a cada uno de éstos se le asignan $\frac{V}{P}$ listas.

* **Estrategias híbridas:** En estos modelos, como el índice 2D introducido en [9] y el 3D en [8], el índice se particiona por términos y documentos al mismo tiempo (2D) e incluso, agregando réplicas (3D). La idea de estas arquitecturas es explotar el *trade-off* entre los costos de comunicación y procesamiento que se requieren para resolver consultas.

Además, los nodos de un motor de búsqueda almacenan su porción del índice en memoria (total o parcialmente), lo que modifica los modelos de costos. Resultados experimentales muestran que es posible obtener mejoras si se incorpora la arquitectura del cluster (cantidad de nodos, procesadores y núcleos) en la optimización.

2. Escalables: Para tratar con el crecimiento en la cantidad de información que generan algunos servicios como los sitios de microblogging y redes sociales, junto con la necesidad de realizar búsquedas en tiempo real, son necesarios algoritmos y estructuras de datos escalables. Los aspectos principales a tener en cuenta en este escenario son la tasa de ingestión de documentos, la disponibilidad inmediata del contenido y el predominio del factor temporal [3] [1].

Para satisfacer estas demandas, resulta indispensable mantener el índice invertido en memoria principal. Dado que este es un recurso limitado, se trata de mantener solamente aquella información que permita alcanzar prestaciones de efectividad razonables (o aceptables) [5].

Un primer aporte del grupo [20] consiste en un conjunto de invalidadores de entradas en el índice invertido. Siguiendo esta línea, se propone el desarrollo de una familia de algoritmos de invalidación y poda selectiva (dinámica) [17] de las entradas en el índice, tanto a nivel del vocabulario como de las *posting lists*. Esto se logra con estrategias que monitorizen la evolución y dinámica del vocabulario.

3. Algoritmos Eficientes: Entre las técnicas más utilizadas para mejorar la performance en motores de búsqueda a gran escala se encuentran las técnicas

de *caching*, las cuales se basan en la idea fundamental de almacenar en una memoria de rápido acceso los ítems que van a volver a aparecer en un futuro cercano, de manera de obtenerlos sin incurrir en acceso a disco.

En una arquitectura típica de un motor de búsqueda se implementan caches para resultados [18], posting lists [28], intersecciones [15] y documentos [22]. Nuestro grupo se enfoca en el problema de las intersecciones. Aquí se propone diseñar políticas de admisión y reemplazo que consideren el costo de ejecutar una consulta [10], y no solamente *hit-ratio*. Esta línea es particularmente interesante en escenarios actuales ya que cuando el índice invertido se encuentra en memoria principal el cache de listas pierde sentido. Por otro lado, integrar diferentes caches permite optimizar el uso de espacio, lo que impacta positivamente en las prestaciones [23].

Otra dirección posible es tratar de optimizar la estrategia de caching incorporando información proveniente de redes sociales. En trabajos previos del grupo se ha mostrado que los temas que son tendencia en redes sociales guardan relación con el aumento de la popularidad de una consulta relacionada al mismo [19] y permiten mejorar la performance del cache. Esta línea de trabajo es prometedora ya que el uso de esta clase de información ha mostrado resultados positivos en otros ámbitos (por ejemplo, para mejorar el rendimiento de CDNs).

b. Grandes Datos en Motores de Búsqueda

Los motores de búsqueda son probablemente uno de los primeros ejemplos del uso de Grandes Datos. Las demandas de recolección de documentos, almacenamiento, análisis, gestión y búsqueda requieren de sofisticados algoritmos que operan sobre arquitecturas paralelas y distribuidas. Además, la información generada por las búsquedas de los usuarios (consultas, clicks, etc.) se convierte en información muy valiosa a partir de la cual es posible encontrar patrones de comportamiento y obtener estadísticas acerca de cómo los usuarios interactúan con los buscadores. Algunos trabajos [11] [12] ya mostraron la potencialidad de estas técnicas.

Esta propuesta global propone optimizar procesos internos de un buscador por lo que se considera que existen oportunidades de optimización que abren nuevos problemas y temas de investigación.

c. Indexación Distribuida

Este es un problema real en una máquina de búsqueda de escala web. Los documentos son procesados de forma distribuida y el resultado final debe ser un índice invertido particionado por algún criterio (como se mencionó anteriormente) que pueda ser implementado en un cluster. En los últimos años, además, se han propuesto nuevas estructuras de datos avanzadas que ofrecen un mejor rendimiento en la recuperación (en algunos contextos), como Block-Max [7] y Treaps [13].

Esta línea de investigación se centra en estudiar, diseñar y evaluar algoritmos de construcción de índices sofisticados como los mencionados utilizando estrategias comunmente utilizadas en el ámbito de Grandes Datos (por ejemplo, sobre el framework Hadoop [26]) y tratar de determinar cómo influyen algunos parámetros como el tamaño de la colección y la arquitectura del cluster a utilizar.

Resultados y objetivos

El objetivo principal en este proyecto es estudiar el problema de las búsquedas a gran escala e incorporar el análisis de datos masivos para mejorar las prestaciones de los sistemas de búsqueda. Para ello, se pretende desarrollar, aplicar, validar y transferir modelos, algoritmos y técnicas que permitan construir herramientas y/o arquitecturas para abordar algunas de las problemáticas relacionadas con las búsquedas en Internet, en diferentes escenarios, donde la masividad, velocidad y variedad de los datos es una característica. Se propone profundizar sobre el estado del arte y proponer nuevos enfoques, en particular:

a) Diseñar estructuras de datos eficientes con base en los modelos recientemente propuestos, para soportar índices invertidos distribuidos (en memoria primaria o secundaria).

b) Utilizar análisis de datos masivos para optimizar los algoritmos de búsquedas, a partir de comprender de mejor manera la dinámica de las consultas y sus costos asociados.

c) Diseñar nuevas técnicas de caching complementándolas con información del análisis de redes sociales para mejorar tanto las políticas de admisión como las de reemplazo.

d) Diseñar y evaluar estrategias de indexación distribuida para estructuras de datos avanzadas

usando frameworks del ecosistema de Grandes Datos.

e) Diseñar arquitecturas para aplicaciones específicas de búsquedas ad-hoc para problemas concretos, donde una solución de propósito general no es la más eficiente.

f) Adaptar y transferir las soluciones a diferentes dominios de aplicación como motores de búsqueda de propósito general, verticales y empresariales, redes sociales y servicios móviles, principalmente.

Formación de Recursos Humanos

Este proyecto brinda un marco para que algunos docentes auxiliares y estudiantes lleven a cabo tareas de investigación y se desarrollen en el ámbito académico. Junto con el doctorado del primer autor hay en finalización una tesis de la maestría en “Exploración de Datos y Descubrimiento de Conocimiento”, DC, FCEyN, Universidad de Buenos Aires.

Actualmente, se están dirigiendo cuatro trabajos finales correspondientes a la Lic. en Sistemas de Información de la Universidad Nacional de Luján en temas relacionados con el proyecto. Además, hay dos pasantes alumnos y se espera dirigir al menos dos estudiantes más por año y presentar dos candidatos a becas de investigación.

Referencias

- [1] N. Asadi, J. Lin, and M. Busch. Dynamic memory allocation policies for postings in real-time twitter search. *CoRR*, abs/1302.5302, 2013.
- [2] R. A. Baeza-Yates and B. A. Ribeiro-Neto. *Modern Information Retrieval - the concepts and technology behind search*, 2nd ed. Pearson Education Ltd., 2011.
- [3] M. Busch, K. Gade, B. Larson, P. Lok, S. Lucienbill, and J. Lin. Earlybird: Real-time search at twitter. In *Proceedings of the 28th International Conference on Data Engineering, ICDE '12*. IEEE Computer Society, 2012.
- [4] B. B. Cambazoglu and R. A. Baeza-Yates. Scalability and efficiency challenges in large-scale web search engines. In *Proceedings of the*

- Eighth ACM International Conference on Web Search and Data Mining, WSDM, 2015.*
- [5] C. Chen, F. Li, B. C. Ooi, and S. Wu. Ti: An efficient indexing mechanism for real-time search on tweets. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*. ACM, 2011.
- [6] B. Croft, D. Metzler, and T. Strohman. *Search Engines: Information Retrieval in Practice*. Addison-Wesley Publishing Company, 1st edition, 2009.
- [7] S. Ding and T. Suel. Faster top-k document retrieval using block-max indexes. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '11*. ACM, 2011.
- [8] E. Feuerstein, V. G. Costa, M. Marín, G. Tolosa, and R. A. Baeza-Yates. 3d inverted index with cache sharing for web search engines. In *18th International Conference, Euro-Par 2012, August 27-31, 2012.*, 2012.
- [9] E. Feuerstein, M. Marín, M. J. Mizrahi, V. G. Costa, and R. A. Baeza-Yates. Two-dimensional distributed inverted files. In *16th International Symposium of String Processing and Information Retrieval, SPIRE'09, August 25-27, 2009*.
- [10] E. Feuerstein and G. Tolosa. Cost-aware intersection caching and processing strategies for in-memory inverted indexes. In *In Proc. of 11th Workshop on Large-scale and Distributed Systems for Information Retrieval, LSDS-IR'14, 2014*.
- [11] Y. Hu, Y. Qian, H. Li, D. Jiang, J. Pei, and Q. Zheng. Mining query subtopics from search log data. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2012.
- [12] P. Kaushik, S. Gaur, and M. Singh. Use of query logs for providing cache support to the search engine. In *International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 2014.
- [13] R. Konow, G. Navarro, C. L. Clarke, and A. López-Ortíz. Faster and smaller inverted indices with treaps. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '13*. ACM, 2013.
- [14] J. Leskovec, A. Rajaraman, and J. D. Ullman. *Mining of massive datasets*. Cambridge University Press, 2014.
- [15] X. Long and T. Suel. Three-level caching for efficient query processing in large web search engines. In *Proceedings of the 14th international conference on World Wide Web*. ACM, 2005.
- [16] A. McAfee and E. Brynjolfsson. Big data: the management revolution. *Harvard business review*, (90), 2012.
- [17] A. Ntoulas and J. Cho. Pruning policies for two-tiered inverted index with correctness guarantee. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2007.
- [18] R. Ozcan, I. S. Altıngövdü, and O. Ulusoy. Cost-aware strategies for query result caching in web search engines. *ACM Trans. Web*, 5(2), May 2011.
- [19] S. Ricci and G. Tolosa. Efecto de los trending topics en el volumen de consultas a motores de búsqueda. In *XVII Congreso Argentino de Ciencias de la Computación, CACIC.*, 2013.
- [20] E. Rissola and G. Tolosa. Inverted index entry invalidation strategy for real time search. In *Proceedings of the XXI Congreso Argentino en Ciencias de la Computación, CACIC '15*, 2015.
- [21] E. E. Schadt, M. D. Linderman, J. Sorenson, L. Lee, and G. P. Nolan. Computational solutions to large-scale data management and analysis. *Nature Reviews Genetics*, 11(9), 2010.
- [22] T. Strohman and W. B. Croft. Efficient document retrieval in main memory. In *SIGIR 2007: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2007.

- [23] G. Tolosa, L. Becchetti, E. Feuerstein, and A. Marchetti-Spaccamela. Performance improvements for search systems using an integrated cache of lists+intersections. In *Proceedings of 21st International Symposium of String Processing and Information Retrieval, SPIRE'14*, 2014.
- [24] G. Tolosa and E. Feuerstein. Using big data analysis to improve cache performance in search engines. In *Proceedings of the Simposio Argentino de GRANdes DATos (1st ed.) at 44 JAIIO - 44th Argentine Conference on Informatics, AGRANDA '15*, 2015.
- [25] A. Trotman and J. Zhang. Future web growth and its consequences for web search architectures. In *CoRR, vol abs/1307.1179, 2013*, 2013.
- [26] T. White. *Hadoop: The Definitive Guide*. O'Reilly Media, Inc., 1st edition, 2009.
- [27] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding. Data mining with big data. *Knowledge and Data Engineering, IEEE Transactions on*, 26(1), 2014.
- [28] J. Zhang, X. Long, and T. Suel. Performance of compressed inverted list caching in search engines. In *Proceedings of the 17th international conference on World Wide Web, WWW '08*. ACM, 2008.

Análisis, interpretación y toma de decisiones estratégicas en la Ciencia de Datos

Mag. María Alejandra Malberti Riveros, Mag. Raúl Oscar Klenzi, Mag. Graciela Elida Beguerí

Instituto de Informática / Departamento de Informática / Facultad de Ciencias Exactas Físicas y Naturales / Universidad Nacional de San Juan
Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", Rivadavia, San Juan,
Teléfonos: 4260353, 4260355 Fax 0264-4234980, Sitio Web: <http://www.exactas.unsj.edu.ar>
e-mail: {amalberti, rauloscarklenzi, grabeda}@gmail.com

Resumen

Se propone abordar el paradigma de Ciencia de Datos con el objetivo de reconocer, analizar y describir el conjunto de estudios y prácticas inherentes a la misma y aplicable a grandes colecciones de datos provenientes de diferentes áreas tales como Educación, Bibliotecología, Astronomía y redes sociales. Estos datos serán accedidos y analizados por medio de herramientas de software libre licencia AGPL como Knime, Weka, R, Rapidminer y módulos específicos de Python, que se ejecuten en diferentes plataformas de hardware secuenciales, paralelos y distribuidos.

Palabras clave: Ciencia de Datos, Aprendizaje Automático, Minería de Datos, Software Libre

Contexto

La revolución de datos se está extendiendo a diferentes ámbitos, lo que conduce a que en la actualidad las personas que trabajan en muy diversas áreas estén obligadas a entender cómo usar los datos. En otras palabras, la proliferación de datos ha creado una demanda de conocimientos, y esta

demanda está inmersa en muchos aspectos de la cultura moderna similar a lo que ocurría en los años 80 y 90 cuando surgió la necesidad de que todas las personas, independiente de la actividad en que se desempeñaran, tenían que aprender a usar computadoras. El rápido desarrollo de las tecnologías de la Información y Computación en los últimos 20 años ha cambiado muchos campos de la ciencia y la ingeniería. Algunas disciplinas se han visto transformadas por causa de los datos, sea por cantidad como por la dinámica de los mismos, llevando esto al desarrollo de métodos dato-intensivos en el área de la ciencia y la ingeniería. De esta manera se comienzan a acuñar nuevas áreas de conocimiento como la Ciencia de Datos (Data Science -DS-) o Ingeniería Dato-intensivos. (Pierson, L. 2015)

Así como las Ciencias de la Computación, Sistemas de Información, Ingenierías en Sistemas y Computación han sido las principales titulaciones que la academia ha brindado en el estudio del área de la Informática, hoy en día la cantidad de conocimiento necesario, derivado de los diferentes campos involucrados para las aplicaciones de Descubrimiento de Conocimiento en Datos (Knowledge Discovery in Data – KDD-) o Descubrimiento de

Conocimiento en Datos Masivos (Knowledge Discovery in Massive Data – KDDM-) lleva a que a nivel de grado se incorporen conocimientos relativos a Ciencia de Datos, o a nivel de posgrado se propongan titulaciones en Ciencia de Datos.

La presente propuesta se encuentra en instancias de evaluación, por parte del CONSEJO DE INVESTIGACIONES CIENTÍFICAS Y TÉCNICAS Y DE CREACIÓN ARTÍSTICA- CICIPCA, y para su desarrollo durante el bienio 2016-2017. En el marco de la misma se pretende dar continuidad al proyecto “Búsqueda de Conocimientos en Datos Masivos” llevado adelante en el bienio 2014-2015. El mencionado proyecto ha permitido a sus integrantes trabajar en una primera aproximación con datos inherentes al área astronómica especialmente brindados por el grupo GAE (Grupo de Astronomía Extragaláctica) de la FCEFyN y relevados desde el Sloan Digital Sky Survey22 (SDSS).

Es idea del grupo para la presente propuesta, profundizar el conocimiento y utilización de algoritmos específicos de aprendizaje de máquina que permitan describir y predecir comportamientos no solamente en datos asociados al área de la astronomía sino también a las áreas pertenecientes a las redes sociales, así como determinación de perfiles de alumnos universitarios, utilizando técnicas de caracterización multidimensional.

Respecto al abordaje de problemas actitudinales se integran a la presente propuesta, participantes del proyecto “El comportamiento académico de los alumnos de primer año de las carreras de informática de la FCEFyN de la UNSJ Estrategias para mejorar su rendimiento” desarrollado en el periodo 2011-2013, Cod. 21/E875 res037/11CS, quienes han realizado experiencias teóricas y de

campo en la difícil tarea de la caracterización de comportamientos de alumnos, detectando una compleja confluencia de factores que tienen que ver con la personalidad, aspectos vocacionales y su situación social, económica y familiar.

Introducción

A partir de la década del 2000 se produjeron grandes transformaciones; las transacciones realizadas en distintas organizaciones comenzaron a registrarse lo que causó que en el 2009 una base de datos corporativa promedio contenía alrededor de cinco petabytes, o 5.000.000 gigabytes de datos. (Un petabyte - PB- equivale a 10^{15} bytes = 1 000 000 000 000 000 de bytes; Un gigabyte -GB- equivalente a 10^9 =1.000.000.000 -mil millones- de bytes). Así mismo, en 1998 Google ya había comenzado a registrar cada búsqueda realizada. En 2004, Facebook comenzó a registrar cada interacción de sus usuarios, y en 2005 YouTube comenzó a mover en todo el mundo grandes cantidades de datos de video. En esos momentos las tecnologías no eran capaces de manejar las enormes cantidades de datos, pero han evolucionado y en la actualidad requieren de diferentes competencias pues la cantidad de datos generados seguirá creciendo enormemente.

A medida que la demanda de datos crece, también lo hará la demanda de capital humano. Las técnicas cuantitativas (estadísticas, minería de datos, predicción, optimización, etc.) representarán para gestores y analistas el know-how al utilizar el análisis de grandes cantidades de datos para tomar decisiones efectivas; estudiantes en ciencias de datos son parte de ese futuro capital humano.

Si bien la mayoría de las carreras en Ciencia de datos son a nivel de posgrado (datos de EEUU), existe en aumento la creencia de que es necesario la creación de carreras a nivel de grado, con intensa fundamentación en ciencia, tecnología, ingeniería y matemáticas <http://www.kdnuggets.com/2015/10/data-science-education-begin.html>

En la actualidad habilidades de aprendizaje de máquina son fuertemente requeridas en el mercado laboral, a raíz de que las empresas tienden cada vez más a construir sistemas de decisión automatizados. Esto conlleva a promover la incorporación en las curriculas de las carreras dependientes del Departamento de Informática, de los saberes requeridos para la formación de habilidades inherentes a un científico de datos.

Se han desarrollado métodos de bases de datos que almacenan y administran petabytes de datos on-line, haciendo a estos seguros y accesibles vía Internet o sistemas distribuidos de cómputo y que pueden a la vez ser analizados por potentes herramientas de minería de datos (Data Mining -DM). Es así como DM, que se puede definir como la extracción de conocimiento no trivial, en grandes cantidades de datos y que favorece la toma de decisiones, adquiere centralidad en el área de la ciencia de datos. DM es la confluencia de múltiples áreas interdisciplinarias como la estadística, el aprendizaje de máquina (Machine Learning -ML-), reconocimiento de patrones, sistemas de bases de datos, recuperación de información, la World Wide Web, técnicas de visualización, entre otras, que aplicadas a diferentes dominios han permitido su reciente y creciente progreso. Con el objetivo de asegurar que el DM siga siendo beneficioso para la ciencia y la ingeniería es importante analizar los desafíos en que el DM habrá de ingresar conforme

evolucionen la ingeniería y la ciencia dato-intensivas. (Kargupta, H.,2008)

La tarea esencial de quien hace ciencia de datos es transformar datos en bruto en conocimiento para la toma de decisiones. El análisis de uso de las bases de datos de una biblioteca hace necesario insertar la ciencia de datos en éstas, de modo que permita a los bibliotecarios aplicar herramientas de software con las que logren transformar datos, recuperar, analizar y graficar información. Es decir, adquieran un conjunto de habilidades para unir los extremos entre la necesidad de información de un usuario y el conocimiento almacenado en los datos.

Los equipos de trabajo e investigación en DS han conformado una gama relativamente amplia de habilidades; el lenguaje y/o las herramientas de software requeridas para cualquier aplicación pueden seleccionarse según los conocimientos y experiencia previa. Para algunas aplicaciones - especialmente la creación de prototipos y desarrollo - es más rápido y ágil que se utilicen herramientas y tipo de datos ya conocidos.

En este sentido se presenta el relevamiento estadístico realizado por KDNuggets.com respecto de las estructuras de datos y áreas de aplicación donde se expande la minería de datos, así como los algoritmos y herramientas de software más utilizadas por la comunidad científico-productiva.

Entre otras herramientas están Orange, Weka, R, RapidMiner, Knime, Hadoop, MapReduce y el lenguaje de programación Python. Estas herramientas son entornos de prueba de algoritmos de aprendizaje de máquina y extracción de conocimiento en datos en diferentes formatos, con excelentes tutoriales, y disponibles para plataformas Windows, Linux y Mac. Así mismo muchas

compañías como Google, IBM, Amazon y Microsoft, con el objetivo de procesar grandes cantidades de datos, han incorporado en sus API (Application Programming Interface) tecnología de aprendizaje automático que no requiere elevada experiencia por parte de los potenciales usuarios.

- Orange: es un desarrollo en lenguaje Python de la Universidad de Liubiana, Eslovenia, desde 1996 y factible de descargar de <http://orange.biolab.si/download/>.
- Weka: Waikato Environment Knowledge Analysis es un desarrollo en Java de la Universidad de Waikato, Hamilton, Nueva Zelanda desde 1993. Se descarga de <http://www.cs.waikato.ac.nz/~ml/weka/>
- R: (librería Rattle): desarrollado en la Universidad de Auckland, Nueva Zelanda, en 1993 y la librería Rattle, creada por el Dr. Graham Williams. www.r-project.org
- RapidMiner: (originalmente YALE) versión: 5.3.15, desarrollado, en Java (plataforma Eclipse) en la Universidad de Dortmund, Alemania, en lenguaje Java, desde 2001 <http://sourceforge.net/projects/rapidminer/>
- Knime: desarrollado en la Universidad de Constanza, Alemania. <http://www.knime.org/downloads/overview>.
- Hadoop: Se caracteriza por su capacidad de acceder y permitir procesar enormes cantidades de datos sobre hardware relativamente económico, además hacer posible el almacenamiento de datos en un sistema de archivos distribuido (Hadoop File System -HDFS-). Particularmente Apache Hadoop es un framework de código libre que soporta aplicaciones distribuidas con datos

masivos, bajo licencia apache v2. Puede ser configurado para trabajar con miles de máquinas y petabytes de información. Hadoop está inspirado en los papers de Google sobre MapReduce y Google File System.

Líneas de investigación, Desarrollo e Innovación

- Analizar y describir el conjunto de estudios y prácticas requeridos en Ciencia de Datos.
- Construir y validar instrumentos tendientes a recabar datos inherentes a los estudiantes.
- Estudiar y analizar diferentes conjuntos de datos masivos a procesar.
- Evaluar herramientas de software libre para arquitecturas secuenciales, paralelas y distribuidas.
- Descubrir conocimiento desde diferentes conjuntos de datos.

Resultados y Objetivos

En el marco del proyecto se pretende analizar y estudiar el paradigma de Ciencia de Datos, desde el tratamiento de datos provenientes de diversas áreas por medio de la aplicación de estrategias de aprendizaje de máquina y minería de datos, sobre arquitecturas secuenciales, paralelas y distribuidas.

Algunos de los propósitos, son:

- Sugerir un conjunto de saberes convenientes para la formación de un científico de datos, para que los mismos sean incorporados, según corresponda, en las currículas de las carreras pertenecientes al Departamento de Informática de la Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de San Juan -FCEF, UNSJ-.

- Realizar aportes a la toma de decisiones con los datos tratados en las diferentes aéreas del saber.

Formación de Recursos Humanos

En el marco de esta propuesta, se están desarrollando dos trabajos finales de licenciatura en las carreras pertenecientes al Departamento de Informática de la -FCEFN, UNSJ-.

También se dirige una tesis de posgrado, que se encuentra en etapa de redacción del informe final, correspondiente a la Maestría en Informática de la Universidad Nacional de La Matanza.

A la vez se prevé la generación de trabajos finales de posgrado, para la Maestría en Informática de la -UNSJ-, así como la tutela de becarios en el área expuesta.

Recientemente se ha creado en dependencias del Instituto de Informática de la -FCEFN, UNSJ- el Laboratorio de Sistemas Inteligentes para la Búsqueda de Conocimientos en Datos Masivos, ámbito de aplicación sustantiva de esta temática.

Referencias

- ANUIES Innovación curricular en instituciones de educación superior. Pautas y procesos para su diseño y gestión. Colección Documentos. Compilación Medina Cuevas, Lourdes; Guzmán Hernández, Laura Leticia. Primera edición, 2011 ISBN 978-607-451-032-4
- Ayllón, Silvia, Merlino, Aldo, Escanés, Gabriel, Variables que influyen en la deserción de estudiantes universitarios de primer año. Construcción de índices de riesgo de abandono Revista Electrónica "Actualidades Investigativas en Educación" [en línea]

- 2011, Disponible en:<<http://www.redalyc.org/articulo.oa?id=44720020005>> ISSN
- Felmer, L. R., Pool, G. M., Fisher, I. R., & Fritz, C. G. (2009). Los estilos epistémicos y tipos de personalidad como factores asociados a la elección de carrera Epistemological styles and types of personality as factors associated in choice of study areas. *Revista de Pedagogía*, 30(86), 115-134.
- Journey, R. (2013). *Agile Data Science: Building Data Analytics Applications with Hadoop*. " O'Reilly Media, Inc."
- Kargupta, H., Han, J., Philip, S. Y., Motwani, R., & Kumar, V. (Eds.). (2008). *Next generation of data mining*. CRC Press.
- Pierson, L., Swanstrom, R., & Anderson, C. (2015). *Data Science for Dummies*. John Wiley & Sons.
- Sarro, L. M., Eyer, L., O'Mullane, W., & De Ridder, J. (Eds.). (2012). *Astrostatistics and Data Mining* (Vol. 2). Springer Science & Business Media.
- Van Der Aalst, W. (2011). *Process mining: discovery, conformance and enhancement of business processes*. Springer Science & Business Media.
- Way, M. J., Scargle, J. D., Ali, K. M., & Srivastava, A. N. (Eds.). (2012). *Advances in machine learning and data mining for astronomy*. CRC Press.
- Williams, G. (2011). *Data mining with Rattle and R: The art of excavating data for knowledge discovery*. Springer Science & Business Media.
- Zhao, Y. (2012). *R and data mining: Examples and case studies*. Academic Press.

Extracción y clasificación automatizada para la Unidad de Vigilancia Tecnológica e Inteligencia Competitiva de la Patagonia

Claudio Delrieux, Damián Barry, Romina Stickar, Luís Ignacio Aita, Juan Manuel Cortez

Laboratorio de Investigación en Informática
Departamento de Informática, Facultad de Ingeniería,
Universidad Nacional de la Patagonia San Juan Bosco.
Puerto Madryn, Chubut, Argentina. +54 280-4472885 – Int. 116
cad@uns.edu.ar, damián_barry@unpata.edu.ar, romistickar@gmail.com, ignacioaita@gmail.com,
juanmanuelcortez@gmail.com

Resumen

El presente trabajo presenta la constitución, desarrollo y actividades realizadas por la Unidad de Vigilancia Tecnológica e Inteligencia Competitiva de la Patagonia. En particular el desarrollo realizado de una herramienta que permite automatizar y homogeneizar las ecuaciones de búsqueda que utilizarán los expertos de la Unidad de Vigilancia.

Palabras clave: big data analytics, extracción, clasificación, Vigilancia Tecnológica, Innobación Abierta.

Contexto

La Unidad de Vigilancia Tecnológica e Inteligencia Competitiva de la Patagonia (VTeIC Patagonia) está conformada por los consorcistas que integran el Parque Tecnológico Puerto Madryn (PTPM) y tiene la finalidad de incorporar a la ciudad de Puerto Madryn una plaza más de la red de Antenas de Vigilancia Tecnológica e Inteligencia Competitiva que estratégicamente están planteadas dentro del Plan Argentina Innovadora 2020, formando parte del Programa Nacional de

Vigilancia Tecnológica e Inteligencia Competitiva (VINTEC) del Ministerio de Ciencia, Tecnología e Innovación Productiva de la República Argentina (MinCyT) [1][2][10].

Las actividades presentes se enmarcan en el proyecto de investigación “Clasificación de Información en BigData mediante la utilización de Técnicas de Inteligencia Artificial y Análisis de Redes Sociales” que el Laboratorio de Investigación en Informática (LINVI) acredita en la Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB).

Introducción

Los territorios deben enfrentar nuevos desafíos para el diseño de estrategias de desarrollo dentro de un contexto de mayor complejidad, incertidumbre y velocidad de cambios, adquirir mayores competencias, adaptarse a las exigencias del mercado y avanzar hacia el desarrollo del territorio[10].

Así, la utilización de las potencialidades endógenas se presenta como la estrategia para lograrlo. Teniendo en cuenta que la difusión de las innovaciones y el conocimiento entre las empresas y la organización de los

sistemas productivos en formas más flexibles, son dos pilares fundamentales para el proceso que mejoran las economías internas de las firmas y favorecen el posicionamiento competitivo de las ciudades y territorios[10].

Puerto Madryn presenta espacios institucionales activos en los que participan representantes de diferentes sectores que avanzan en experiencias concretas de articulación y cuyo aprendizaje aportan las bases para la creación de una oficina de VTelC. A la vez, esta unidad permitirá enriquecer tales espacios inter-institucionales

En la actualidad existe la necesidad de administrar grandes volúmenes de información no estructurada. Este incremento, debido a la enorme producción de información digital, ya sea desde la perspectiva de información producida en internet como así también la enorme cantidad de información producida por las empresas y organismos que en su gran mayoría, por no ser administradas correctamente, se termina dejando sin uso en algún repositorio de información. Normalmente esta información termina siendo eliminada sin evaluar correctamente su utilidad por parte de una comunidad.

La producción y obtención de información ha pasado a ser uno de los grandes activos de las organizaciones, ya sean públicas, mixtas o privadas. En este sentido el desarrollo y estudio de la generación, administración, explotación, interpretación y clasificación de información se ha convertido en un desafío tecnológico y científico a nivel mundial. Para poder abordarlo, no sólo se requiere del soporte de científicos y tecnólogos en el área de la informática sino además de la integración con investigadores y expertos de distintas áreas vinculadas con las actividades que

se desean analizar y comprender, donde a través de la conformación de equipos multidisciplinarios generen verdadero valor a la información circundante.[1][2]

Un Sistema de Vigilancia e Inteligencia consta de 7 (siete) fases preliminares: planificación, identificación de necesidades, búsqueda de información y herramientas, monitoreo y validación, tratamiento y análisis, difusión y protección y evaluación y seguimiento conformando lo que se llama el Ciclo de VTelC.

- Fase 0: Estructura Organizativa, planificación de actividades de VT, proyección de productos y servicios de VTelE, recursos físicos y humanos, etc.
- Fase 1: Identificación de necesidades e interpretación del sector - árbol tecnológico, fuentes de información, definición de palabras claves y términos técnicos, recopilación documental, distribución geográfica.
- Fase 2: Búsqueda de información, herramientas y generación de ecuaciones de búsqueda.
- Fase 3: Monitoreo y validación de la información.
- Fase 4: Tratamiento y análisis de información.
- Fase 5: Difusión y protección de la información.
- Fase 6: Evaluación, seguimiento y actualización del proceso de VeIE y presentación de Informe Final

Herramienta de Vigilancia automatizada

Como hemos expresado entendemos que no es posible pensar implementar un tratamiento manual para la gestión de conocimiento realizado por expertos para la implementación de las fases 2, 3 y 4

planteadas en la sección anterior, es por esto que la Unidad de VTeIC de la Patagonia ha diseñado y desarrollado una herramienta integral que permita realizar la extracción de información tanto científica, de patentes como comercial y de mercado de forma automatizada y homogénea para su tratamiento.

Para realizar este producto se establecieron 3 etapas:

1. Extracción y pre-clasificación de información.
2. Clasificación avanzada de información y construcción de espacios semánticos y taxonómicos mediante técnicas de inteligencia artificial.
3. Explotación de la información clasificada. Automatización de los informes y reportes de la Unidad de Vigilancia Tecnológica e Inteligencia Competitiva.

Líneas de Investigación, Desarrollo e Innovación

Las líneas actuales de investigación se enmarcan dentro de las 3 etapas propuestas en la sección anterior. Para ello se llevan adelante las siguientes líneas de investigación.

- Bases de datos no estructuradas NoSQL. Incluyendo conceptos de Extract, Transform and Load (ETL) para la recuperación y homogeneización de información heterogénea[8][9].
- Técnicas de recuperación de información (information retrieval)[11][12]. Dentro de esta línea se incluyen entre otras temáticas:
 - Lematización.
 - Reconocimiento de entidades nombradas.
 - Extracción de relaciones.

- Técnicas de análisis de redes sociales para la categorización, evaluación y valuación de la red de autores, temáticas y países.
- Técnicas de machine learning y datamining orientado al análisis de textos.
- Técnicas de visualización inteligente de información.

Resultados y Objetivos

Motor de extracción de información

Para la extracción de contenido público de bibliotecas digitales y portales de patentes se desarrolló un motor de extracción de contenido público de internet denominado CrawlingExtractor. El mismo está basado en un trabajo anterior para extraer noticias de orden público georreferenciadas según su contenido denominado “ZCrawler”[4], el mismo se compone de un lenguaje de extracción que le permite a los usuarios mediante una especificación formal comenzar a extraer información siguiendo múltiples criterios y permitiendo la clasificación de esta información según a quien se está evaluando.

Para ello se desarrollaron servicios de ejemplo para las API de Sciencedirect[5] Springer[6], Scopus[7] y para medios digitales se construyeron con RSS estándar.

Adicionalmente se adaptaron todos los extractores ya que las definiciones de API y/o RSS no accedían a información complementaria de los artículos publicados como por ejemplo el perfil de los autores y las citas bibliográficas.

Arquitectura de Extractores

El éxito de la herramienta se basa en la posibilidad de estandarizar en un sólo formato y lenguaje todos los sitios y plataformas de extracción de información, evitando de esta forma la necesidad que

los expertos aprendan muchos lenguajes para elaborar las ecuaciones de búsqueda. Además el esfuerzo de trabajar con formatos heterogéneos de información y estandarizarlos permite luego, independientemente de la fuente unificar la forma de consultar la información.

Para ello es necesario que se programen los distintos formatos de extracción para cada fuente. Si bien el mismo requiere de un esfuerzo por parte de programadores, una vez desarrollado la misma fuente sirve para múltiples consultas.

Para almacenar las extracciones se utiliza una base de datos NoSQL orientada a la construcción de índices invertidos denominada Apache Solr[8], basada en el motor de índices de Apache Lucene[9].

Actualmente la herramienta cuenta con 4 extractores definidos y asociado a cada uno ecuaciones de búsqueda sobre el sector Aluminio en particular en las temáticas relacionadas a los materiales de colada de aluminio y el sector de Pesca y Acuicultura particularmente orientada a la producción de alimento wakame a partir del alga undaria pinnatifida.

Próximos Objetivos

1. Mejorar el esquema de extracción de información en lo referente a lematización, stopwords y reconocimiento de atributos.
2. Extraer información relevante de publicaciones, autores y citas a una base de datos orientada a grafos para realizar análisis de redes sociales (ARS) especialmente en determinar clústers de información relacionada.
3. Incorporar clasificación automática de tópicos de los

documentos mediante técnicas de machine learning.

Formación de Recursos Humanos

- Actualmente 2 alumnos de la carrera de grado realizan sus tesinas en el marco del presente proyecto. Cada uno enmarcado en las líneas de investigación comentadas.
- Como parte de las actividades del proyecto de Investigación, actualmente se está desarrollando el Curso de posgrado “Introducción a la ingeniería de Ontologías”

Referencias

1. Edgar Morin. El método iii. el conocimiento del conocimiento. Madrid: Cátedra, 2:24, 1988.
2. Edgar Morin and Marcelo Pakman. Introducción al pensamiento complejo. Gedisa Barcelona, 1994.
3. Wayan Vota, Rajendra Singh, Siddhartha Raja, Jude Genilo, Shamsul Islam, Marium Akther, and Rohan Samarajiva. Digital government: building a 21st century platform to better serve the american people. 2012.
4. Barry Damián, Aita Luis Ignacio, and Cortez Juan Manuel. Zcrawler: Extracción, clasificación y publicación de información pública desde su perspectiva geográfica. JAIIO 2014.
5. Elsevier B.V. Sciencedirect.
6. Springer Science+Business Media. Springer.
7. Elsevier B.V. Scopus.

8. Apache Foundation. Apache solr reference guide.
9. Apache Foundation. Apache Lucene Core.
10. Jorge A Sabato. Ensayos en campera. Juárez, 1979.
11. Charu C. Aggarwal • ChengXiang Zhai. Mining Text Data. Springer Science+Business Media. 2012.
12. Katariina Nyberg. Document Classification Using Machine Learning and Ontologies. Espoo, January 31, 2011.
13. Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining Practical Machine Learning Tools and Techniques. Third Edition. Elsevier Inc. 2011

ARQUITECTURA GENÉRICA PARA EL ALMACENAMIENTO DE DATOS BIOMÉTRICOS

Silvia Ruiz, Ernesto Miranda, Graciela Etchart, Carlos Alvez, Marcelo Benedetto, Juan José Aguirre, Santiago Delfin, Mauro Herlein

Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos

Av. Tavella 1424, Concordia, Entre Ríos - CP 3200

silruiz@fcad.uner.edu.ar, emiranda@fcad.uner.edu.ar, getchart@fcad.uner.edu.ar,
caralv@fcad.uner.edu.ar, marben@fcad.uner.edu.ar, juaagu@fcad.uner.edu.ar,
toti.delfin@gmail.com, herlein.mauro@gmail.com

Resumen

El objetivo de este trabajo es el desarrollo de una arquitectura genérica utilizando la tecnología de base de datos objeto relacional (BDOR), conforme a normas internacionales, para la identificación de las personas por medio de reconocimiento biométrico. La necesidad de contar con modelos basados en estándares, es permitir la interoperabilidad entre sistemas, para de esa manera, facilitar la búsqueda y el intercambio de datos. En el presente proyecto, se ha propuesto una arquitectura genérica para Iris, basada en el registro Tipo 17 del estándar ANSI/NIST-ITL 1-2011. La idea es extender esta arquitectura para el rasgo de voz con el objeto de incrementar la fiabilidad del sistema utilizando el reconocimiento multimodal.

Palabras clave: modelo objeto-relacional base de datos biométricas, interoperabilidad.

Contexto

Este trabajo se da en el marco del Proyecto PID 07/G044 “*Gestión de datos biométricos en base de datos objeto - relacionales*”, que da continuidad al Proyecto PID 07/G035 “*Identificación de*

personas mediante Sistemas Biométricos. Estudio de factibilidad y su implementación en organismos estatales” [1][2][3][4][5].

El objetivo general del presente proyecto es desarrollar sistemas para la identificación de personas mediante el reconocimiento de iris y voz que permitan la interoperabilidad entre organismos acorde a normas internacionales. Este trabajo, se enfoca principalmente en el desarrollo de arquitecturas de bases de datos para la representación e intercambio de datos biométricos. Como caso de aplicación, se tratarán imágenes de iris para el reconocimiento biométrico [6][7].

Introducción

La relevancia del uso de tecnologías basadas en biometrías se encuentra en su aplicación como mecanismo de control que a través de una serie de medidas de características específicas permite el reconocimiento de personas, para superar los problemas de vulnerabilidad e inconvenientes (gastos en su creación, control, administración, posibles extravíos, olvidos, etc.) del sistema de autenticación por contraseñas, tarjetas de acceso, entre otros.

En el campo de la biometría, los estándares definen, entre otros aspectos,

los formatos de intercambio para el almacenamiento y la transmisión de los datos biométricos. Dichos formatos son importantes para aplicaciones en las que el ancho de banda es limitado y/o cuando el registro del rasgo biométrico debe ser almacenado en dispositivos con limitada capacidad.

En los últimos años, la Argentina ha demostrado ser un país con potencial en la región para implementar y desarrollar soluciones de biometría tanto para el sector público como privado [8][9].

Los organismos públicos que emplean esta tecnología necesitan consultar las bases de datos biométricas de manera eficiente e intercambiar datos con otros organismos asociados. De allí la necesidad de contar con modelos de datos y con estándares que permitan la interoperabilidad entre sistemas y así facilitar la búsqueda e intercambio de datos en bases biométricas. El ANSI/NIST-ITL 1-2011 [10], contempla este tipo de aspectos (entre otros). Esta norma biométrica publicada en noviembre de 2011, es el estándar más utilizado por entes estatales, y sobre el mismo se viene trabajando en el grupo de investigación del proyecto [3]. En diciembre de 2013, una actualización del estándar incorporó la especificación para el registro Tipo 11 referido al tratamiento de la voz como rasgo biométrico, lo cual es importante porque es uno de los rasgos sobre el que se trabajará en el presente proyecto.

La elección de la voz, como otro rasgo biométrico a fusionar con el iris, se debe que, además de cumplir propiedades como: universalidad, distintividad, estabilidad, evaluabilidad, no-intrusivo, etc., es un rasgo de fácil adquisición. Puede ser adquirida de una manera muy sencilla sin métodos invasivos, ni dispositivos especializados. Esto hace de la voz, un rasgo biométrico ideal para la arquitectura propuesta.

Actualmente, en el grupo de investigación se ha propuesto una arquitectura genérica para la gestión de imágenes de iris [11].

En la misma (Figura 1), las tareas de generación, transmisión y recepción de registros de Tipo 17 del estándar se encuentran contempladas dentro del sistema de intercambio de imagen del iris. Este sistema tiene un módulo que se encarga de la generación y transmisión de los registros necesarios para una transacción ANSI / NIST ITL 1-2011 y otro módulo para la recepción de estos registros a partir de organismos asociados. Tanto la estructura de registro de Tipo 17 como los metadatos para el reconocimiento del iris, son estructuras complejas. Esto puede causar algunos problemas cuando se trabaja con el modelo de datos relacional, debido a las limitaciones impuestas por el mismo. Por lo tanto, aquí se ha optado por la tecnología objeto-relacional (OR) [12].

Por otra parte, el intercambio de imágenes requiere en muchos casos transacciones web, con lo cual el tamaño de las mismas es importante. Un aspecto a tener especialmente en cuenta es la calidad de representación de las imágenes. En este sentido, los estándares de intercambio de datos también exigen que los formatos de compresión utilicen métodos que cumplan determinados requisitos. El estándar ANSI/NIST-ITL 1-2011, para imágenes de iris, sólo admite JPEG 2000 (Estándar ISO/IEC 15444:2004) [13] y PNG (Estándar ISO/IEC 15948:2004) [14]. Este aspecto también es contemplado en la arquitectura propuesta.

El estándar internacional (JPEG 2000) ISO/IEC 15444:2004 es un sistema de codificación de imágenes, que permite una compresión, transmisión y almacenamiento eficiente de imágenes fijas y de secuencias de imágenes. Con este estándar es posible adecuar la secuencia de bits que representa la información de la imagen comprimida a las características del canal de transmisión, del dispositivo de almacenamiento, o del dispositivo de visualización, independientemente del

tamaño, el número de componentes, y la precisión de la imagen original.

El estándar internacional ISO/IEC 15948: 2004 PNG (Gráficos de Red Portables) especifica un flujo de datos y un formato de archivo asociado para una imagen sin pérdidas, comprimida, portátil y que puede ser transmitida a través de Internet. Entre las principales características del formato PNG se encuentra la portabilidad. El formato es flexible a aceptar agregados y extensiones futuras sin comprometer el intercambio de los flujos de datos PNG.

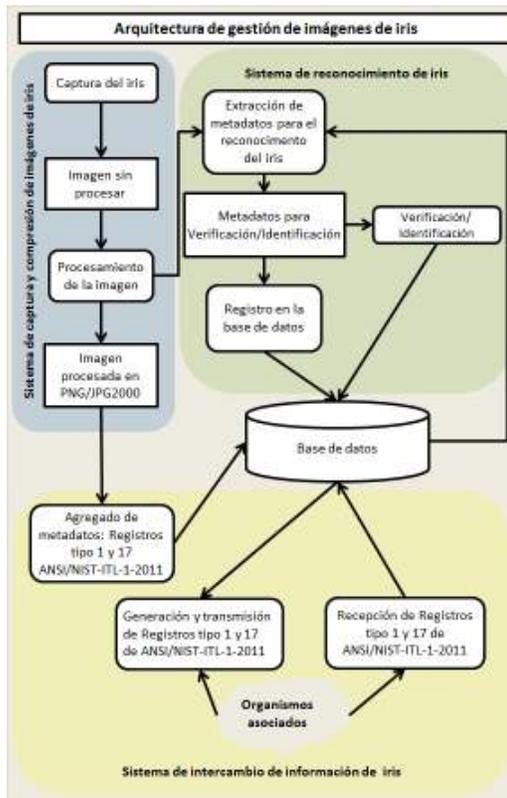


Figura 1: Arquitectura para gestión de imágenes de iris

Líneas de Investigación y Desarrollo

Una de las líneas más importantes, consideradas en el proyecto, consiste en la construcción de modelos para la representación de datos relacionados con los datos biométricos, de manera que los mismos puedan compartirse, consultarse, recuperarse y compararse de manera

simple y eficiente. En particular, el reconocimiento del iris es ampliamente utilizado para fines de seguridad. Asimismo, como se mencionó con anterioridad, especialmente en organismos gubernamentales, es muy importante considerar el nivel de estandarización alcanzado, sobre todo, en lo referente al formato de intercambio de datos y a una arquitectura de base de datos que facilite el almacenamiento y la recuperación de los datos biométricos. En nuestro caso, por su capacidad de adaptación a diferentes dominios, se utilizará el modelo de base de datos objeto relacional.

Otra línea de investigación vigente está orientada a sistemas que combinan diferentes rasgos biométricos o bases de datos biométricas multimodales. Estos sistemas son más precisos y seguros dado que superan algunas limitaciones de los sistemas que utilizan un único rasgo biométrico. Las limitaciones de un rasgo pueden ser: no universalidad, ruidos en los datos, suplantación de identidad, entre otros [15]. A pesar de que el tiempo de procesamiento, el hardware requerido y los requisitos de almacenamiento en bases de datos multimodales pueden ser mayores a los empleados por las bases de datos unimodales, las ventajas de aquellas hacen que se prefiera su utilización. [16] Por lo antes expuesto, se extenderá la arquitectura propuesta para el tratamiento del rasgo de voz (registro Tipo 11 del estándar ANSI/NIST-ITL 1-2011).

Resultados y Objetivos

El objetivo general del Proyecto 07/G044 es desarrollar sistemas para la identificación de personas mediante el reconocimiento de iris y voz que permitan la interoperabilidad entre organismos. El objetivo específico en que se está trabajando y que se presenta en este artículo, es el desarrollo de una arquitectura de base de datos que facilite

el almacenamiento y la recuperación de los datos biométricos.

En esta etapa del proyecto se encuentra en desarrollo una arquitectura genérica utilizando la tecnología de las BDOR, de acuerdo con las normas internacionales, para la identificación de las personas por medio de reconocimiento de iris. Para dicha arquitectura se propuso un modelo expresado en un diagrama de clase UML en el que se definen los tipos de datos de dominio a utilizar en la arquitectura.

La arquitectura presentada se adaptará para el soporte de datos biométricos de voz. Esto deberá contemplar, al igual que con el iris, lo especificado en el estándar ANSI/NIST-ITL 1-2011, los métodos para extracción de características y el almacenamiento de las secuencias de los vectores a utilizar para el reconocimiento [17].

En etapas posteriores se trabajará en la implementación de los métodos que gestionen los tipos específicos para los datos biométricos crudos, los tipos de datos para la codificación de los mismos, y todos los metadatos necesarios para generar los registros de transacciones ANSI/NIST-ITL1-2011. Además, se estudiarán alternativas para la creación de índices de dominio que permitan mejorar los tiempos de respuestas en los procesos de reconocimiento.

Formación de Recursos Humanos

La estructura del equipo de investigación está formada por el director y co-director, cuatro integrantes docentes. El Director del proyecto dirige la tesis de Maestría en Sistemas de Información (MSI) de la Facultad de Ciencias de la Administración de la UNER de un integrante. Además, dos integrantes y la Co-Directora se encuentran cursando la MSI, donde realizarán sus trabajos de investigación en el área del proyecto.

Además, el proyecto cuenta con un becario de Iniciación en la Investigación cuyas tareas están relacionadas con la captura, registración y almacenamiento de datos biométricos y un integrante alumno, de la Licenciatura en Sistemas, que colabora en la implementación de la arquitectura propuesta.

Referencias

1. Carlos E. Alvez, Marcelo G. Benedetto. "Los Sistemas Biométricos y su Factibilidad de Aplicación en los Organismos Estatales". XII Workshop de Investigadores en Ciencias de la Computación (WICC 2010), El Calafate, Santa Cruz Argentina, 5 y 6 de Mayo de 2010, Universidad Nacional de la Patagonia San Juan Bosco. Páginas 247-251.
2. Graciela Etchart, Lucas Luna, Carlos Leal, Marcelo Benedetto, Carlos Alvez. Sistemas de reconocimiento biométricos, importancia del uso de estándares en entes estatales. CGIV - XIII Workshop de Investigadores en Ciencias de la Computación (WICC 2011), 5 y 6 de Mayo de 2011. Universidad Nacional de Rosario. Rosario - Argentina. Páginas 339-343.
3. Graciela Etchart, Lucas Luna, Rafael Leal, Marcelo Benedetto, Carlos Alvez. "Sistema adecuado a estándares de reconocimiento de personas mediante el iris". CGIV - XIV Workshop de Investigadores en Ciencias de la Computación (WICC 2012), 25 y 26 de Abril de 2012. Universidad Nacional de Misiones. Posadas - Argentina. Páginas 321-325.
4. Graciela Etchart, Carlos Alvez, Marcelo Benedetto "Gestión de Datos Biométricos en Bases de Datos Objeto-Relacionales" XV Workshop de Investigadores en Ciencias de la Computación (WICC 2013). pp. 97-101. Paraná Entre Ríos.

5. Silvia Ruiz, Graciela Etchart, Carlos Alvez, Ernesto Miranda, Marcelo Benedetto, Juan José Aguirre "Representación e interoperabilidad de imágenes biométricas" XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015). Salta.
6. Daugman, J.: High condence visual recognition of persons by a test of statistical independence. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(11), 1148-1161 (November 1993).
7. J. Daugman and C. Downing, "Effect of severe image compression on iris recognition performance," IEEE Transactions on Information Forensics and Security, vol. 3, no. 1, p. 52-61, 2008.
8. Casal Gabriel, Revolva Mercedes. Biometrías. Herramientas para la Identidad y la Seguridad Pública. Jefatura de Gabinete de Ministros. Presidencia de la Nación. Noviembre de 2010.
9. Julio Fuoco, Tendencias Biométricas, desafíos y oportunidades. En Biometrías 2. Jefatura de Gabinete de Ministros. Presidencia de la Nación. Octubre de 2011
10. Wing B. ANSI/NIST-ITL 1-2011. Update: 2013. Information Technology: American National Standard for Information Systems Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information. December, 2013.
11. Carlos Alvez, Graciela Etchart, Silvia Ruiz, Ernesto Miranda, Juan Aguirre, Marcelo Benedetto, Santiago Delfin "Iris Information Management in Object-Relational Databases". XXI Congreso Argentino de Ciencias de la Computación (Junín, 2015), pp. 741-750. Junin, Argentina. Octubre 2015.
12. Melton Jim, "(ISO-ANSI Working Draft) Foundation (SQL/ Foundation)", ISO/IEC 9075-2:2003 (E), United States of America (ANSI), (2003).
13. ISO/IEC 15444-1:2004 Information Technology-JPEG 2000 Image Coding System: Core Coding System, Second Edition, 2004-09-15.
14. ISO/IEC 15948:2004 Information Technology-Computer Graphics and Image Processing-Portable Netware Graphics (PNG): Functional Specification, First Edition, 2004-03-01.
15. Alvez C., Benedetto M., Berón G., Etchart G., Luna L. y Leal C. Desarrollo de un sistema multi-biométrico mediante reconocimiento de iris y voz, adecuado a estándares, para su aplicación en organismos públicos. SIE 2011 - Simposio de Informática en el Estado. Córdoba, 31 de Agosto, 01 y 02 de Septiembre de 2011. 40° JAIIO. Páginas: 206 - 220.
16. M. Ghayoumi, "A review of multimodal biometric systems: Fusion methods and their applications," Computer and Information Science (ICIS), 2015 IEEE/ACIS 14th International Conference on, Las Vegas, NV, 2015, pp. 131-136
17. Reynolds, D. A., Campbell, W., Gleason, T. T., Quillen, C., Sturim, D., Torres-Carrasquillo, P., Adami, A., 2005. The 2004 MIT Lincoln Laboratory speaker recognition system. In: Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, ICASSP. Vol. 1. pp. 177-180.

Técnicas de Clasificación aplicadas al rendimiento académico

Myriam Herrera¹, María Inés Lund², Susana Beatriz Ruiz¹, Estela Liliana Torres¹, Lilian Adriana Mallea³, María Gema Romagnano²

¹Departamento de Informática, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de San Juan

²Instituto de Informática, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de San Juan

³Departamento de Matemática, Facultad de Filosofía, Humanidades y Artes, Universidad Nacional de San Juan

mherrera, mlund, mromagnano@{info.unsj.edu.ar}

Resumen

En muchas investigaciones se tiene necesidad de identificar cuáles son las características que diferencian unos grupos de sujetos u objetos respecto de otros, para así poder realizar predicciones. El análisis de conglomerados y el análisis discriminante, son técnicas que algunos autores ubican entre las más potentes para aplicar en investigaciones sociales, permiten clasificar sujetos u objetos a partir de características similares.

Estas dos técnicas se pueden diferenciar por la manera de extraer conocimiento útil escondido en esos datos. El Análisis Discriminante cuenta con grupos de datos conocidos, con observaciones de unidades de pertenencia desconocida inicialmente y tiene que ser determinada a través del análisis de los datos. Este tipo de problemas de clasificación es referido como reconocimiento de patrones asistido o aprendizaje supervisado; en terminología estadística cae bajo el título de Análisis Discriminante.

Por otro lado, hay problemas de clasificación donde los grupos son desconocidos a priori y el principal

propósito del análisis es determinar los grupos a partir de los propios datos, de modo que las unidades dentro del mismo grupo sean, en algún sentido, más similares u homogéneas que aquellas que pertenecen a grupos diferentes. Este tipo de problema de clasificación es referido como reconocimiento de patrón no supervisado o conocimiento sin guía, y, en terminología estadística cae bajo el título de Análisis de Conglomerados.

En este proyecto se aplicarán ambas técnicas o una combinación de ellas o una nueva técnica para analizar lo que llamamos rendimiento académico universitario. Se puede afirmar que, en general, un indicador directo de la calidad de la enseñanza es el rendimiento académico, medido a través del nivel alcanzado por los estudiantes. Vista la importancia del tema en este proyecto se determinarán las principales variables que influyen en el rendimiento como así también tipologías básicas de grupos, obtenidos de los alumnos universitarios tanto de la Facultad de Ciencias Exactas como de los alumnos de matemática de la Facultad de Filosofía de la UNSJ.

Palabras clave: Clasificación, Rendimiento, Calidad Universitaria

Contexto

Este proyecto se encuentra en un estado inicial, ya que recientemente ha sido aprobado por evaluación externa, es de carácter bi-anual y financiado por la UNSJ. Se encuentra inserto en el marco de las líneas de investigación de los Gabinetes Estadística e Ingeniería de Software del Instituto de Informática de la FCFN de la UNSJ.

Además se vincula a cátedras de las carreras de Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas de Información, que se dictan en la Institución.

Introducción

Los datos se han convertido en un recurso crítico en muchas organizaciones y por lo tanto, el acceso eficiente a estos, el compartirlos, extraer información de los mismos y hacer uso de la información extraída se transforma en una imperiosa necesidad. La necesidad de comprender conjuntos grandes y complejos de datos, ricos en información, es común a todos los campos de los negocios, ciencia, ingeniería. [1, 3, 11, 14, 15, 16]

Como resultado hay muchos esfuerzos no sólo para integrar varias fuentes de datos dispersos a través de sitios diferentes, sino también es importante la información extraída de esas bases de datos en bajo la forma de patrones y tendencias.

El proyecto se basa en dos principales ejes de investigación y aplicación:

Data Mining [3, 11] analiza conjuntos de datos para encontrar relaciones y resúmenes de datos útiles para el propietario de los datos. Estas relaciones y resúmenes derivados a través del

ejercicio del Data Mining se refieren a modelos y patrones.

La aplicación automatizada de algoritmos de minería de datos permite detectar fácilmente patrones en los datos. Los algoritmos de minería de datos se clasifican en dos grandes categorías: supervisados o predictivos y no supervisados o de descubrimiento del conocimiento.

El Reconocimiento de Patrones tiene como objetivo la **clasificación** de objetos dentro de un número de categorías o clases [7, 8, 12]. Dependiendo de la aplicación estos objetos pueden ser imágenes, señales o cualquier tipo de medidas que necesitan ser clasificadas. Esas medidas se llaman patrones.

Las medidas usadas para la clasificación de objetos o patrones son conocidas como características. El conjunto de todas las características forman el vector que identifica únicamente a un patrón (objeto).

Líneas de Investigación, Desarrollo e Innovación

En muchas de las investigaciones, independientemente del área de conocimiento, es habitual tener la necesidad de identificar cuáles son las características que diferencian unos grupos de sujetos u objetos respecto de otros, para así poder realizar predicciones futuras. Tanto el análisis de conglomerados como el análisis discriminante son técnicas que nos permiten clasificar sujetos u objetos a partir de características similares. La diferencia fundamental entre ambas pruebas es el momento del establecimiento de los grupos. En el análisis discriminante (AD) [2, 10, 13] el investigador conoce a priori a qué grupo pertenece cada sujeto u objeto; en

cambio, en el análisis de conglomerados los grupos o clúster se determinan y configuran a posteriori, es decir, una vez estudiadas y analizadas las agrupaciones.

El análisis discriminante es la prueba estadística apropiada para seleccionar qué variables independientes o predictivas permiten diferenciar grupos y cuántas de estas variables son necesarias para alcanzar la mejor clasificación posible. Además permite cuantificar su poder de discriminación en la relación de pertenencia de un sujeto u objeto a un grupo u otro. Por ello esta técnica es considerada, además de una prueba de clasificación, una prueba de dependencia. De hecho, su propósito es similar al análisis de regresión logística; la diferencia radica en que solo admite variables cuantitativas.

En el presente proyecto utilizaremos estas técnicas en el ámbito educativo como es el estudio del rendimiento estudiantil y la identificación de las variables que mejor lo predicen, a partir de las calificaciones de materias que clasifican al alumnado en grupos bien diferenciados.

Mediante un análisis discriminante se puede establecer el poder explicativo y discriminatorio de las características que diferencian a los alumnos según su rendimiento. Además del rendimiento se tendrán en cuenta en el estudio una serie de variables independientes como, por ejemplo, variables de carácter socioeconómico, variables académicas referentes a la preparación en el nivel secundario y variables actitudinales en relación con la variable dependiente que clasifica a los sujetos según el rendimiento obtenido [4, 9].

Según las características analizadas, a través de la descripción del grado de relación existente entre el conjunto de variables, se puede encontrar la frontera que separa los grupos.

Como resultado obtendremos una regla de clasificación que podrá ser utilizada en el pronóstico de adscripción al grupo de rendimiento establecido para nuevos estudiantes..

Resultados y Objetivos

Este proyecto es bi-anual y está iniciándose en 2016, con lo cual no tenemos actualmente resultados.

Objetivo General:

- Determinación de una función discriminante que explique la influencia de un conjunto de variables en el rendimiento académico de alumnos universitarios.

Objetivos Específicos:

- Determinar las Variables influyentes en el rendimiento académico de alumnos universitarios.
- Identificar qué variables tienen mayor poder de discriminación y de predicción en la clasificación de sujetos
- Determinar si existen diferencias significativas entre los “perfiles” de un conjunto de variables de dos o más grupos definidos a priori.
- Establecer un procedimiento, función discriminante, para clasificar a un individuo a partir de los valores de un conjunto de variables.
- Evaluar la exactitud de la clasificación mediante la regla de decisión que asigne un objeto nuevo a uno de los grupos prefijados con un cierto grado de riesgo

Formación de Recursos Humanos

El equipo de investigación está formado por docentes investigadores de dos facultades de la UNSJ, y las unidades de observación serán los datos de

alumnos de Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas de Información del Departamento de Informática y Licenciatura en Matemática, del Departamento de Matemática.

Se espera sumar alumnos tesis de grado y posgrado (maestría y doctorado), interesados en esta línea de investigación.

Referencias

1. ATO, M.; LÓPEZ, J.A. (1996): Análisis estadístico para datos categóricos. Madrid: Editorial Síntesis
2. BENZÉCRI, Jean Paul (1976): L'Analyse des données, T.I La taxonomie T.II L'Analyse des correspondances. Dunod. París.
3. CHENGKAI Li, «CSE4334/5334 Data Mining», University of Texas at Arlington, 2015.
4. DIAZ, M., PEIO, A., ARIAS, J., ESCUDERO, T., RODRIGUEZ, S., VIDAL, G. J. (2002). Evaluación del Rendimiento Académico en la Enseñanza Superior. Comparación de resultados entre alumnos procedentes de la LOGSE y del COU. En: Revista de Investigación Educativa, 2 (20), 357-383.
5. DIDAY, Edwin, y LECHEVALLIER, Yves (1991): Symbolic Numeric data analysis and learning, Versailles, September 18-20. INRIA, Nova Science Publishers Inc. New York.
6. DIDAY, Edwin (1992): Analyse des données et classification automatique numérique et symbolique. Seminario Internacional de Estadística en Euskadi. Volumen 27. EUSTAT, Euskal. Estatistika Erakundea/ Instituto Vasco de Estadística.
7. EUSTAT, Vitoria-Gasteiz. (1997): Análisis de datos simbólicos. Ed. IRICE, Rosario.
8. FERNÁNDEZ AGUIRRE, Karmele: IV International Meeting of Multidimensional Data Analysis (NGUS'97), Bilbao, September 10-12, 1997. Universidad del País Vasco, Bilbao.
9. GARBANZO VARGAS, Giuselle María (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la Educación Superior Pública. Revista Educación, 31(1), 43-63, ISSN: 0376-7082.
10. GONZALEZ LOPEZ, Ignacio (2004). Realización de un Análisis discriminante explicativo del rendimiento académico en la Universidad. Revista Investigación Operativa, vol 22, nº1, 43-59. Universidad de Córdoba.
11. HAN, Jiawei y KAMBER, Micheline, Data Mining. Concepts and Techniques, 2.a ed. Morgan Kaufmann, 2006.
12. LEBART, Ludovic; MORINEAU, Alain, y PIRON, Marie (1995): Statistique exploratoire multidimensionnelle. Dunod. París.
13. TORRADO FONSECA, Mercedes; BERLONGA-SILVENTE, Vanesa (2013). Análisis Discriminante mediante SPSS. REIRE (Revista d'Innovació i Recerca en Educació).
14. WEB. «Data mining made faster: New method eases analysis of "multidimensional" information», ScienceDaily. <https://www.sciencedaily.com/releases/2010/07/100722075013.htm>. [Accedido: 15-feb-2016].

15. WEB. «Big Data, for better or worse: 90% of world's data generated over last two years», *ScienceDaily*. <https://www.sciencedaily.com/releases/2013/05/130522085217.htm>. [Accedido: 15-feb-2016].
16. WEB «The Four V's of Big Data», *IBM Big Data & Analytics Hub*. <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>. [Accedido: 15-feb-2016].

Evolución del Modelo Multidimensional en un DataWarehouse para pacientes diabéticos

E. Mangia¹, D. Omar¹, J.P. Madonni¹, M. E. Llorente¹, J. Besso¹, A. Sigura^{1,2}, A. J. Hadad^{1,2},

B. Drozdowicz^{1,2}

¹Facultad Ciencia y Tecnología, Universidad Autónoma de Entre Ríos

²Facultad Ingeniería, Universidad Nacional de Entre Ríos

Ruta 11, Oro Verde, Entre Ríos, Argentina

mellorente@arnet.com.ar, bdrozdo@santafe-conicet.gov.ar

Resumen

En el presente trabajo se propone una nueva aplicación de la estrategia de diseño iterativo para el Modelo Multidimensional de un DataWarehouse propuesto en etapas anteriores y que respondía al primer caso de uso definido. Este caso de uso estaba orientado a responder consultas sobre la condición de la enfermedad de un paciente. La propuesta actual es mostrar la evolución del modelo aplicando los requerimientos definidos en el segundo caso de uso, orientado a evaluar la enfermedad en grupo de pacientes. Los Casos de Uso (CU) que fueron definidos inicialmente en el trabajo de investigación, representan los diferentes requerimientos planteados por usuarios. Esta forma iterativa de diseño permite la incorporación de nuevos requerimientos desde los distintos ámbitos de interés en el tema.

Palabras clave: Datawarehouse, Pacientes Diabéticos, Casos de Uso, Estructuras de Datos, Modelo Multidimensional, ETL.

Contexto

El presente trabajo se inserta en un Proyecto de Investigación Plurianual (PIDP) denominado “*Sistema de Soporte a la Toma de Decisiones basado en datawarehouse para pacientes diabéticos*”. Dicho proyecto es desarrollado en

la Facultad de Ciencia y Tecnología de la Universidad Autónoma de Entre Ríos (FCYT - UADER).

Introducción

De acuerdo a la propuesta de una Metodología Iterativa de diseño, en cada ciclo, el objetivo es poder resolver requerimientos propuestos [1]. Esto permite ampliar el modelo y permitir dar respuesta a nuevos requerimientos. Fundamentalmente cuando en el problema inicial planteado, no se conocen la totalidad de los requerimientos de los distintos usuarios [2]. El diseño obtenido anteriormente aplicando el concepto del *Modelo Multidimensional* [3], estaba basado en los dos Casos de Uso orientados al nivel de paciente [2]. El objetivo de este trabajo es presentar el análisis necesario para lograr la evolución del mencionado diseño frente a los requerimientos de un tercer *Caso de Uso* [2]. Los requerimientos definidos en los *Casos de Uso*, están divididos en niveles [2]. La evolución se realiza utilizando el Caso de Uso orientado al segundo nivel que corresponde al *Grupo de Pacientes*.

Líneas de investigación y desarrollo

De las líneas de investigación descriptas para

este proyecto en el trabajo presentado en el WICC 2012 [4], en esta propuesta se analizaron las siguientes líneas:

1. Estructuras de datos representativas del dominio de análisis.
2. Métodos ETL para fuentes de información de referencia.
5. Granularidad y dimensiones del modelo.
6. Metadatos y reglas de negocios.

Resultados y Objetivos

El Modelo Multidimensional del DW para enfermos diabéticos que es utilizado para la implementación del prototipo es el representado en la Fig. 2. Este modelo resulta de los requerimientos propuestos en el primer Caso de Uso (Fig. 1)

Recuperación de Imágenes
Objetivo: Recuperar dos imágenes realizadas al mismo paciente en tiempos diferentes.
Contexto: La identidad del médico y sus privilegios de acceso han sido validados
Actores: Médico
Recursos: El PACS y el DataWarehouse están accesibles
Episodios (*):
a. El médico ingresa el nombre del paciente (o nro de registro) → Listar los estudios según el paciente
b. El médico elige el tipo de estudio → Listar los estudios según el paciente y el tipo de estudio, indicando el momento de realización de cada uno
c. El médico selecciona una o más imágenes → Transferir las imágenes a la computadora cliente
(*): Acción del Actor → Responsabilidades del sistema

Figura 1. Casos de Uso orientados al primer nivel

Sobre este primer diseño también se realizó un trabajo de análisis y definición de las ETL's que permiten el correspondiente procesamiento de información [5].

A continuación, se analizó el diseño de la Fig.2 de acuerdo a los requerimientos propuestos en el Caso de Uso orientado al segundo nivel de usuarios (Fig.3).

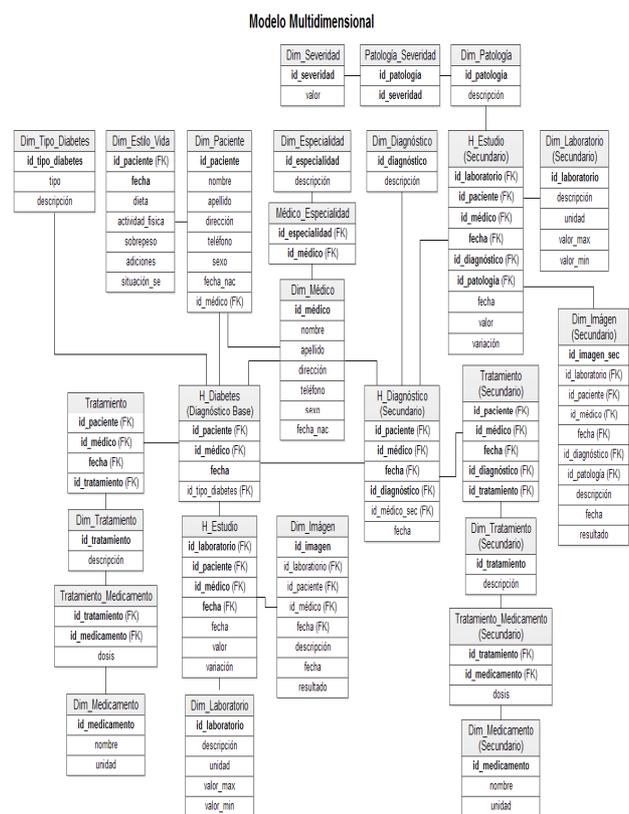


Figura 2. Modelo Multidimensional propuesto.

Analizando los requisitos detallados en ambos Casos de Uso, se observa necesaria la incorporación de información que permita comparar resultados de un análisis con los valores de referencia. Se decide incluir en la fuente de información aquella que permita realizar esta comparación, la cual permita comparar el resultado de estudios de pacientes y establecer el desvío del mismo. De esta forma sería factible relacionar el resultado de un estudio con una patología determinada.

Este requerimiento responde al Caso de Uso correspondiente a la Fig.3 (1er columna).

Teniendo en cuenta a las nuevas necesidades, se propone el nuevo modelo multidimensional mostrado en la Fig. 4, que es una evolución respecto del anterior.

Comparación de Imágenes	Análisis de un Grupo de Pacientes
Objetivo: Comparar imágenes para detectar diferencias en alguna evolución de la patología y mostrar al usuario las imágenes crudas y los resultados de la comparación	Objetivo: Estadística para un grupo de pacientes que compartan condiciones patológicas equivalentes.
Contexto: La identidad del médico y sus privilegios de acceso han sido validados	Contexto: La identidad del médico y sus privilegios de acceso han sido validados
Actores: Médico	Actores: Médico
Recursos: El PACS y el DataWarehouse están accesibles	Recursos: El PACS y el DataWarehouse están accesibles
Episodios (*): a. El médico ingresa el nombre del paciente (o nro de registro) → Listar los estudios según el paciente. b. El médico elige el tipo de estudio → Listar los estudios según el paciente y el tipo de estudio, indicando el momento de realización de cada uno c. El médico selecciona dos o más imágenes → Transferir las imágenes a la computadora cliente d – El médico elige dos imágenes para comparar → Ejecutar el método de comparación, Generar Imagen resultado de la comparación y Transferir imagen resultado a la computadora cliente	Episodios (*): a. El médico ingresa / selecciona una patología nombrada. → Listar los estudios asociados a dicha patología indicando la severidad diabética b. El médico elige el nivel de severidad que desea observar → Listar todos los estudios según los pacientes pertenecientes a este nivel de severidad. c. El médico selecciona las variables a analizar (edad, peso, sexo, etc) en los estudios → Ejecutar el método estadístico estándar y graficar histograma para representar media, mediana, moda, entre otras. d. El médico selecciona el subgrupo según el nivel de severidad → Transferir Resultado a la computadora cliente.

Figura 3. Caso de Uso orientados al segundo nivel

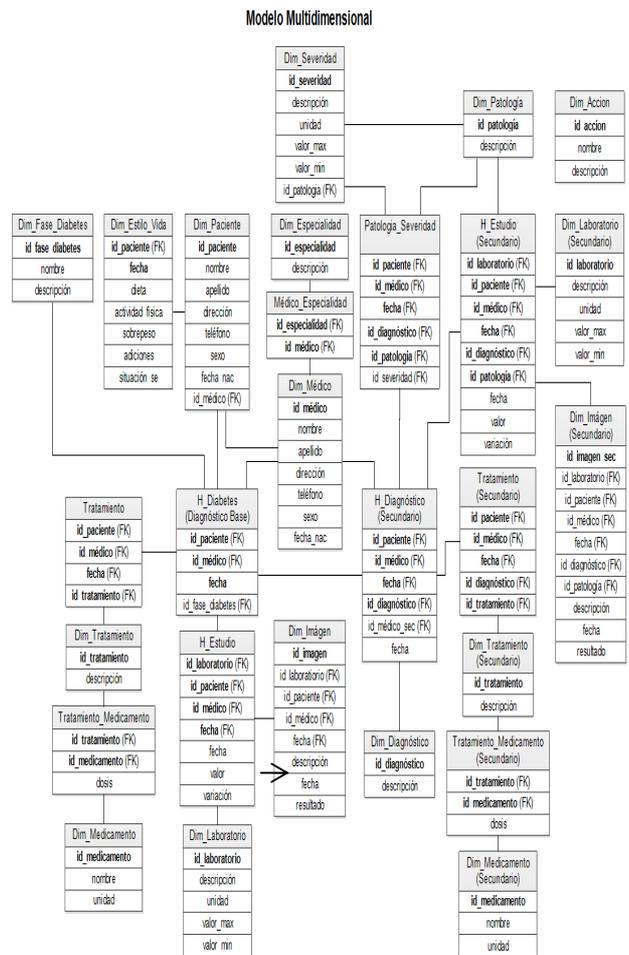


Fig. 4 Modelo Multidimensional para el segundo Ciclo de Vida

Se aplicaron las siguientes modificaciones:

- Se agregó la dimensión Dim_Patología asociada a la H_Estudio
- Esta nueva dimensión se relaciona con la severidad a través de Patología_Severidad que permite definir la dimensión Dim_Severidad
- También se agregó la dimensión Dim_Imagen relacionada con la tabla de hechos H_Estudio (Secundario) entendiendo que también estudios secundarios pueden incluir imágenes como resultados.

Respecto del Caso de Uso representada en la 2da.columna de la Fig.3, las modificaciones

son:

- Se agregaron los campos unidad, valor_max, valor_min a la dimensión Dim_Severidad con el objetivo de almacenar los valores de referencia de la misma.
- Se modificó la cardinalidad “muchos a muchos” entre Dim_Severidad y Dim_Patologia por la cardinalidad “1 a muchos” (una patología puede tener n severidades y una severidad pertenece solo a una patología). Esto es debido a que diferentes severidades pueden tener la misma descripción (por ejemplo: Normal) pero pueden representar diferentes valores de referencia dependiendo de la patología.
- Se modificó Dim_Tipo_Diabetes por Dim_Fase_Diabetes, para determinar el estado de evolución de la diabetes.

De esta manera, se ha obtenido un modelo que responde a diferentes requerimientos y fundamentalmente aplicando el proceso iterativo propuesto inicialmente [6].

Sobre este diseño, no solamente se analizó la factibilidad de dar respuesta a los requerimientos propuestos en los Casos de Uso, sino que además se modificaron las fuentes de información tomadas como origen. Este análisis es fundamental para poder redefinir los procesos de ETL's necesarios para la carga de información.

Se propone como camino a seguir, la aplicación de una nueva iteración aplicando requerimientos orientados a un tercer nivel de usuarios (instituciones y organizaciones médicas) [2].

Esta necesidad, impone la definición de un nuevo Caso de Uso que respete las necesidades de este nuevo nivel de usuarios. A tal efecto, se propone el Caso de Uso de la Fig.5, el cual tiene por objetivo conocer si la cantidad de médicos que actualmente trabaja en la institución es suficiente para atender todas las

consultas de los pacientes.

Análisis de los pacientes atendidos por una institución de salud	
Objetivo	Determinar si la cantidad de médicos de la institución es suficiente para atender todas las consultas de pacientes
Contexto	La identidad del usuario y sus privilegios de acceso han sido validados
Actores	Dirección médica y autoridades institucionales
Recurso	El DW está accesible
Episodios	<i>Acción del actor ⇒ Responsabilidades del sistema El usuario selecciona las variables que pueden combinarse con la presente consulta (especialidad, patología, edad, sexo, peso, estatura, etc) ⇒ El sistema ejecuta la consulta junto con las variables seleccionadas y genera un reporte para el usuario.</i>

Figura 5. Caso de Uso orientados al tercer nivel

Para poder dar respuesta a este nuevo CU, es necesario disponer de información relacionada a los turnos de todos los profesionales médicos, como así también del tiempo de cada consulta de paciente. Con dicha información es posible calcular el promedio de pacientes diarios que atiende cada médico y de esta forma poder determinar si la cantidad de médicos es suficiente para cubrir todas las consultas de los paciente.

El modelo multidimensional propuesto en la Fig. 4 no almacena la información requerida para dar soporte CU anteriormente definido.

Es trabajo futuro realizará una nueva evolución para aplicar las modificaciones necesarias al modelo multidimensional y de esta forma dar respuesta al CU referido al tercer nivel de usuarios.

Formación de Recursos Humanos

El equipo de trabajo está conformado por especialistas del área de sistemas de información, inteligencia artificial y bioingeniería. Integrantes del equipo tienen formación de postgrado tanto en el área de sistemas de información como en el área biomédica, así como también experiencia en el ámbito profesional en lo que refiere al desarrollo de sistemas.

Se incorporan tres alumnos becarios de la carrera Licenciatura en Sistema de la Facultad de Ciencia y Tecnología de la UADER

Referencias

[1] M.G. Piattini, J.A. Calvo-Manzano, J. Cervera, L.Fernandez. Análisis y Diseño de Aplicaciones informáticas de Gestión. Una perspectiva de Ingeniería del Software. 2007, ISBN 978-84-7897-776-5. RA-MA Editorial

[2] M. E. Llorente, Aldo Daniel Sigura, Javier Besso, Alejandro Hadad, Bartolomé Drozdowicz. “Proceso de Diseño basado en Casos de Uso para un Datawarehouse Clínico”. CACIC 2012

[3] M. E. Llorente, Aldo Daniel Sigura, Javier Besso, Ernesto Mangia, Alejandro Hadad, Maria Manzini, Noralí Quijada y Bartolomé Drozdowicz. “Propuesta de un Modelo Multidimensional para Datawarehouse sobre pacientes diabéticos” WICCC 2014, Ushuaia, Tierra del Fuego.

[4] M. E. Llorente, Aldo Daniel Sigura, Alejandro Hadad, Bartolomé Drozdowicz. “Sistema de soporte a la toma de decisiones basado en datawarehouse para pacientes

diabéticos.” WICC 2012

[5]. M. E. Llorente, A. Sigura, J. Besso, E. Mangia, A. J. Hadad, M. Mancini, N. Quijada, B. Drozdowicz. “Aplicación de prototipo para el Modelo Multidimensional de un Datawarehouse sobre pacientes diabéticos”. WICC 2015, Ushuaia, Tierra del Fuego.

[6] E.Mangia, O.Denis, M.E.Llorente, J.Besso, A.Sigura, B. Drozdowicz, A.Hadad. “Proceso iterativo de diseño del modelo multidimensional de soporte a datos clínicos de pacientes diabéticos”. Ventana Informática. Universidad de Manizales.2015.ISSN 0123-9678

METODOLOGÍA PARA EVALUACIÓN DE IMPACTO DE MIGRACIÓN ENTRE VERSIONES DE BASES DE DATOS NoSQL

Giovanni Daián Róttoli^{1,2}, Juan Zaffaroni¹, Marcelo López Nocera¹, Ma. Florencia Pollo-Cattaneo^{1,2}

¹ Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS). Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina.

² Ingeniería en Sistemas de Información, Facultad Regional Concepción del Uruguay. Universidad Tecnológica Nacional. Argentina
{gd.rottoli, flo.pollo}@gmail.com

RESUMEN

Día a día, el crecimiento sostenido de tecnologías como Internet y la proliferación de herramientas como las redes sociales disparan el surgimiento de necesidades en materia de almacenamiento y procesamiento de datos. Las distintas tecnologías NoSQL existentes, soluciones en desarrollo de las problemáticas asociadas al fenómeno Big Data, se actualizan constantemente para dar respuesta a esta amplia variedad de requerimientos. Sin embargo, estas nuevas versiones de los sistemas de base de datos NoSQL pueden dar lugar a cambios que impacten de manera diversa sobre los distintos procesos que hacen uso de los mismos. Por dicho motivo, en este trabajo se presenta la propuesta de elaboración de una metodología para la evaluación del impacto de la realización de migraciones entre distintas versiones de tecnologías NoSQL, a fin de otorgar herramientas que permitan a los usuarios evaluar alternativas y optimizar la toma de decisiones en lo que se refiere a la implementación de tales actualizaciones.

Palabras clave: NoSQL, Migración, Análisis de Impacto

CONTEXTO

La presente investigación se desarrolla dentro de una línea de trabajo incipiente del Grupo de Estudio de Metodologías para Ingeniería en Software y Sistemas de Información (GEMIS),

conformado por un equipo de docentes y alumnos de la Universidad Tecnológica Nacional (UTN), pertenecientes a las Regionales Buenos Aires y Concepción del Uruguay. El grupo GEMIS, se halla abocado a la búsqueda de la sistematización de cuerpos de conocimientos y promoción sobre el campo de la Ingeniería en Sistemas de Información y la Ingeniería en Software, sus aplicaciones y abordajes metodológicos en todo tipo de escenarios (convencionales y no convencionales).

En el marco de la UTN, el equipo de investigadores ha trabajado en forma conjunta desde los Departamentos de Ingeniería en Sistemas de Información de ambas Regionales y desde la Escuela de Posgrado de la Regional Buenos Aires, integrando entre sus miembros a docentes de grado y de posgrado, articulando los resultados de investigaciones con el desarrollo de Trabajos Finales de Carrera, Trabajos Finales de Especialidad y Tesis de Maestría.

1. INTRODUCCIÓN

De manera habitual las organizaciones mutan para adecuarse al impacto de los cambios del medio circundante. Dichas mutaciones involucran desde el aspecto humano hasta las

tecnologías de soporte a la toma de decisiones [Moya, 2008].

Las bases de datos, son una de las herramientas que deben adecuarse continuamente a los distintos y cambiantes requerimientos de los usuarios [Lopez, 2012]. Por tal motivo, en la última década, han surgido un conjunto de alternativas a las bases de datos tradicionales, que luchan contra las problemáticas acarreadas por el fenómeno del Big Data. Las denominadas bases de datos NoSQL han sido desarrolladas para brindar esquemas flexibles en el trabajo con datos no estructurados, alta eficiencia ante consultas de datos, y posibilidades de escalar horizontalmente, sacrificando cuestiones como la consistencia de los datos u otras características deseables en las bases de datos SQL [Abramova et al., 2014, Nayak et al., 2014, Strauch et al., 2011].

Bajo esta propuesta se establecen cuatro grandes tipos de tecnologías NoSQL [Arora & Aggarwal, 2013; Bugiotti & Cabibbo, 2013; Hecht & Jablonski, 2011, Moniruzzaman & Hossain, 2013; Nayak et al., 2013; Strauch et al., 2011]:

- **Documentales**, las cuales permiten almacenar información en forma de documentos sin un esquema predefinido, usualmente haciendo uso de archivos JSON o XML referenciados por claves únicas. Ejemplos de ello son: MongoDB [Mongodb.org, 2016], CouchDB [Couchdb.apache.org, 2016] y Riak [Basho, 2016].
- **Clave-Valor**, que resultan muy simples y eficientes. Asimilándose a mapas y diccionarios, almacenan información en dos partes, una parte “valor” (que

almacena información), y una parte “clave” (que referencia a dicho valor). Los registros no pueden relacionarse entre sí, sino en la lógica de la aplicación que las utiliza. Ejemplos de bases de datos Clave-Valor son Redis [Redis.io, 2016] y Voldemort [Project Voldemort, 2016].

- **Columnares**, permiten referenciar a partir de una clave un conjunto de valores o columnas que pueden, a su vez, agruparse en familias de columnas, admitiendo un nivel de organización de la información, superior al de las bases de datos clave-valor. Sin embargo, al igual que éstas, persisten en la imposibilidad de relacionar entidades. De esta manera operan bases de datos como Cassandra [Cassandra.apache.org, 2016] y Hbase [Hbase.apache.org, 2016].
- **Gráficas**, acceden al almacenamiento de la información en forma de grafos, a través de nodos y relaciones entre ellos. Estas bases de datos son muy eficientes para realizar consultas entre entidades altamente relacionadas (como relaciones de amistad entre entidades personas, o productos relacionados con operaciones de compra) pues se sirven de las aplicaciones de la Teoría de Grafos. La utilización de bases de datos gráficas puede observarse en Neo4J [Neo4j Graph Database, 2016].

Como se puede apreciar, cada tipo de base de datos NoSQL posee un conjunto de características que deben tenerse en cuenta a la hora de seleccionar la más adecuada para cada organización [Pollo-Cattaneo et al., 2014]. La

importancia de la valoración de los factores distintivos de cada tecnología se evidencia aún más, si se ponen en perspectiva el auge y la democratización del acceso a Internet y las telecomunicaciones, la impronta que el avance de éstas confiere al surgimiento de nuevas necesidades y la fuerza con la que imponen el desarrollo e implementación de funcionalidades novedosas que den respuesta a tales requerimientos emergentes; demandas que el mercado informático ha de cubrir, poniendo a disposición las más variadas características en cada nueva versión de la tecnología en cuestión [Stanier, 2012].

Sin embargo, en ocasiones, al elegir operar con versiones actualizadas, se sacrifican utilidades que los usuarios valoran por lo que es altamente conveniente que en la elección para la utilización de un motor de base de datos NoSQL, se considere de importancia no solo el tipo de base de datos, sino también, el impacto que ha de tener la migración desde una instancia del software hacia otra, desde el punto de vista de las implicaciones en los procesos que hacen uso de las mismas [Pollo-Cattaneo et al., 2014].

Ante esta situación, se hace necesaria una metodología que permita la comparación de versiones de bases de datos NoSQL a fin de poder evaluar el impacto de migración entre dichas versiones.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Durante 2015, a pedido de una organización del sector privado, el grupo GEMIS realizó un análisis comparativo de dos versiones de MongoDB₇ (base de datos NoSQL de tipo Documental) en un ambiente empresarial, a fin de evaluar las características de ambas, el

impacto que provocan sobre los procesos productivos relacionados y la migración de la tecnología utilizada hacia una versión más actualizada de la misma [Zaffaroni et al., 2016].

Frente a tal escenario, se evidenció la necesidad de contar con una serie de pasos sistematizados a efectos de replicar la evaluación efectuada, sobre cualquier tipo de base de datos no relacional, y obtener precisiones respecto de las cualidades más relevantes de las tecnologías implicadas, y de un posible impacto sobre los procesos productivos que se ponen en juego en las instancias de actualización.

En este contexto, el objetivo que se persigue es el de desarrollar una metodología de trabajo que pueda aplicarse entre versiones de cualquier tipo bases de datos NoSQL y que permita a sus usuarios acceder a herramientas para la selección eficiente de la más adecuada. (La metodología propuesta será evaluada ante casos concretos a fin de asegurar su validez).

Para cumplir con el objetivo establecido, se seguirán los siguientes pasos:

1. Realizar una investigación documental exploratoria sobre las características relevantes de las bases de datos NoSQL.
2. Identificar casos y datos para las futuras pruebas.
3. Seleccionar las condiciones sobre las cuales se ejecutarán los casos de prueba y validación.
4. Desarrollar una propuesta inicial de la metodología.
5. Estudiar el comportamiento del modelo utilizando los casos de prueba y validación sobre los datos propuestos.

6. Analizar los resultados obtenidos y realizar modificaciones hasta contar con una versión estable.

3. RESULTADOS Y OBJETIVOS ESPERADOS

Mediante la ejecución del procedimiento descrito en el apartado anterior, se busca desarrollar una metodología de trabajo para la comparación de versiones de bases de datos NoSQL a fin de evaluar los posibles impactos de migración entre ambas tecnologías.

Dicha metodología permitirá conocer si ciertas características presentes en determinadas versiones presentan cambios y el tipo de cambio en cuestión (actualización o eliminación de la característica, entre otros). De esta manera, se podrá comparar si un cambio de versión en el sistema de bases de datos NoSQL afectaría al sistema del cual forma parte.

El proyecto propuesto se inscribe en una línea de investigación dentro del ámbito educativo universitario mediante el estudio y análisis de diferentes tecnologías de Bases de Datos relacionales y no relacionales. Con ello, se pretende continuar y ampliar los trabajos desarrollados por GEMIS en el área de Bases de Datos, brindando conocimiento valioso para volcar en las aulas de grado y posgrado y, en una segunda instancia, realizando difusión de los mismos dentro del ámbito científico.

4. FORMACIÓN DE RECURSOS HUMANOS

Este proyecto aspira, tanto a la obtención de nuevos conocimientos y a la formación de especialistas en el análisis de la adopción de procesos vinculados con la Ingeniería de Software, como a la promoción del ascenso de

los recursos humanos implicados, dentro del escalafón de la carrera de investigadores. El grupo de trabajo está integrado por dos investigadores formados, un investigador en formación y un becario doctoral. Además, uno de los miembros se encuentra en desarrollo de su Trabajo Final de Especialidad.

Asimismo, en el marco de este proyecto de investigación se prevé la radicación de una Tesis de Maestría en Ingeniería en Sistemas de Información y su articulación con trabajos finales de la carrera de grado de Ingeniería en Sistemas de Información.

5. REFERENCIAS

- Abramova, V., Bernardino, J., & Furtado, P. (2014). Experimental evaluation of NoSQL databases. *International Journal of Database Management Systems*, 6(3), 1.
- Arora, R., & Aggarwal, R. R. (2013). Modeling and Querying Data in MongoDB. *International Journal of Scientific and Engineering Research*, 4(7).
- Basho, (2016). Riak KV. Retrieved 29 February 2016, from <http://basho.com/products/riak-kv/>
- Bugiotti, F., & Cabibbo, L. (2013). A Comparison of Data Models and APIs of NoSQL Datastores. Dipartimento di Ingegneria della Università di Roma.
- Cassandra.apache.org. (2016). The Apache Cassandra Project. Retrieved 29 February 2016, from <http://cassandra.apache.org/>
- Couchdb.apache.org. (2016). Apache CouchDB. Retrieved 29 February 2016, from <http://couchdb.apache.org/>
- Hbase.apache.org. (2016). Apache HBase – Apache HBase™ Home. Retrieved 29 February 2016, from <https://hbase.apache.org/>
- Hecht, R., & Jablonski, S. (2011). NoSQL evaluation: A use case oriented survey. 2011 International Conference on Cloud and Service Computing. 336-341.
- Lopez, D. (2012). Análisis de las posibilidades de uso de Big Data en las organizaciones. Universidad de Cantabria, Santander, España.
- Mongodb.org. (2016). MongoDB. Retrieved 29 February 2016, from <https://www.mongodb.org/>
- Moniruzzaman, A. B. M., & Hossain, S. A. (2013). Nosql database: New era of databases for big data analytics-classification, characteristics and comparison. arXiv preprint arXiv:1307.0191.

- Moya J. (2008). *Management Democrático*. Cataluña: PreMya Consultores.
- Nayak, A., Poriya, A., & Poojary, D. (2013). Type of NOSQL databases and its comparison with relational databases. *International Journal of Applied Information Systems*, 5(4), 16-19.
- Neo4j Graph Database. (2016). Neo4j: The World's Leading Graph Database. Retrieved 29 February 2016, from <http://neo4j.com/>
- Pollo-Cattaneo, M. F., Nocera, M. L., & Rottoli, G. D. (2014). Rendimiento de tecnologías NoSQL sobre cantidades masivas de datos. *Cuaderno Activa*, (6), 11-17. ISSN: 2027 – 8101.
- Project Voldemort,. (2016). Project Voldemort: A Distributed Database. Retrieved 29 February 2016, from <http://www.project-voldemort.com/voldemort/>
- Redis.io,. (2016). Redis. Retrieved 29 February 2016, from <http://redis.io/>
- Stanier, C. (2012). Introducing nosql into the database curriculum. In 10th International Workshop on the Teaching, Learning and Assessment of Databases (p. 61).
- Strauch, C., Sites, U. L. S., & Kriha, W. (2011). NoSQL databases. Lecture Notes, Stuttgart Media University.
- Zaffaroni, J., Róttoli, G.D., López Nocera, M., Pollo-Cattaneo, M.F., (2016). Reporte Técnico: Estudio Comparativo de características de versiones de la base de datos NoSQL Documental MongoDB. Retrieved 02 March 2016, from <https://sistemas.frba.utn.edu.ar/grupogemis/Trabajos/Reportes%20Tecnicos/GEMIS-TD-2016-03-TR-2016-03.pdf>

Personalización de la Educación a través de la creación de Perfiles dinámicos de los alumnos

Javier Díaz¹, Laura Lanzarini², María Emilia Charnelli¹,

¹Laboratorio de Investigación en Nuevas Tecnologías Informáticas.
Facultad de Informática. Universidad Nacional de La Plata
javierd@info.unlp.edu.ar,
mcharnelli@linti.unlp.edu.ar

²Instituto de Investigación en Informática LIDI (III-LIDI).
Facultad de Informática. Universidad Nacional de La Plata
laural@lidi.info.unlp.edu.ar

Resumen

La línea de investigación actual continúa con el estudio y desarrollo en el área de la Minería de Datos Educativa. Se ha puesto énfasis en el análisis y aplicación de diferentes técnicas no supervisadas y de técnicas de visualización de datos masivos con el objetivo de identificar las características más relevantes de los alumnos de la Facultad de Informática de la UNLP. A su vez, se espera poder contribuir en el área educativa a través de la determinación de perfiles dinámicos de los alumnos de la Facultad de Informática en lo que se refiere a su interacción con recursos educativos de acceso libre y su interacción con las redes sociales. Dichos perfiles podrán ser utilizados para caracterizar su comportamiento actual y asistirlo en forma automática a fin de que pueda alcanzar el comportamiento esperado.

Palabras clave: analítica del aprendizaje, minería de datos educativa, sistemas de recomendación

Contexto

Este trabajo de investigación se encuadra en el Proyecto de Incentivos acreditado “Internet del futuro: Ciudades Digitales Inclusivas, Innovadoras y Sustentables, IoT, Ciberseguridad, Espacios de Aprendizaje del Futuro” desarrollado en el Laboratorio de Investigación en Nuevas Tecnologías Informáticas, LINTI.

Introducción

La Analítica del Aprendizaje (Learning Analytics) es un campo de investigación emergente que analiza información referida a personas que están aprendiendo y sus contextos, con el propósito de comprender y optimizar el aprendizaje y los entornos en los que dicho aprendizaje sucede.

En los últimos años, las instituciones educativas se han embarcado en su propia exploración de grandes conjuntos de datos para mejorar los índices de retención y proporcionar una experiencia personalizada y de mayor calidad para los estudiantes. La aplicación de técnicas de

Minería de Datos en el ámbito educativo ha permitido caracterizar a los distintos actores que intervienen en los procesos de enseñanza y aprendizaje [1].

A medida que el aprendizaje se traslada a la Web, crece la cantidad de datos que pueden utilizarse para mejorar y personalizar este proceso.

A continuación se describen las tareas realizadas recientemente.

Análisis de relevancia de atributos

A partir de las líneas de investigación presentadas en WICC 2015 [2], se continuó trabajando sobre los datos académicos y personales de los alumnos de la Facultad de Informática de la UNLP con el objetivo de mejorar la caracterización de los alumnos que pierden la regularidad.

La información utilizada provino del sistema SIU-Guaraní e involucró a 5268 alumnos, comprendidos entre los años 2002 y 2012.

Para la representación de los datos del problema se tuvo en cuenta el trabajo realizado en [3]. Dicha representación, si bien permite incluir los distintos aspectos que caracterizan los motivos de pérdida o no de regularidad por parte del alumno, dan como resultado una cantidad de atributos o variables elevado dificultando la identificación de patrones o relaciones existentes. Por lo tanto, fue preciso aplicar técnicas de Selección de Características y Visualización para encontrar la información más relevante.

A partir de los atributos seleccionados se construyeron tres tipos de modelos diferentes que permitieron clasificar a los alumnos en regulares y no regulares utilizando los siguientes métodos como el C4.5, PART y un multiperceptrón entrenado con el algoritmo de backpropagation. De los modelos obtenidos se pudo afirmar que los

atributos seleccionados son adecuados para predecir la condición de no regularidad de un estudiante. También dejan de manifiesto la no incidencia de la situación laboral actual de los alumnos en lo que se refiere a su rendimiento académico. Los resultados de esta investigación fueron publicados en [4].

También se estudiaron y aplicaron técnicas de aprendizaje no supervisado para evaluar la performance de los alumnos a lo largo de los primeros cinco años de la carrera. Se aplicaron técnicas de clustering para obtener los grupos de alumnos con un desempeño similar en la carrera. Los resultados de esta investigación fueron publicados en [5].

En lo que se refiere al uso de técnicas de visualización como herramienta para seleccionar atributos relevantes consideramos que, en educación, resultan de suma utilidad ya que permiten a los docentes, independientemente del área en que se desarrollen, hacer uso de las mismas para analizar a sus alumnos sin necesidad de contar con conocimientos específicos de la minería de datos.

En esta dirección, se analizaron técnicas de visualización para representar la información de los alumnos de la Facultad de Informática. Se utilizó una visualización de conjuntos paralelos para identificar tendencias en las respuestas [4]. Uno de los resultados obtenidos muestra que de los alumnos egresados, la mayoría de ellos en 5to año sólo tenían aprobado el 50% de la carrera. Así también se realizaron diagramas de desviación sobre los clústeres obtenidos. Como resultado se observó que se produce un punto de inflexión en el segundo año. A partir de ese momento una gran cantidad de alumnos detienen su progreso en la carrera.

Así también, se utilizaron técnicas de visualización para obtener los atributos

más representativos para construir un modelo de clasificación que permite describir y caracterizar a los alumnos según su condición de regularidad. Los resultados de este trabajo fueron publicados en [5].

Minería de Datos Masivos

Además de la información personal y académica obtenida del sistema de gestión y los entornos de enseñanza y aprendizaje, resulta de interés obtener información adicional sobre la interacción de los alumnos con las redes sociales [6]. Actualmente se está realizando un trabajo que analiza las relaciones existentes entre las publicaciones realizadas por los alumnos de la Facultad de Informática de la UNLP en las redes sociales. En este caso, las publicaciones son de texto libre y no están estructuradas por lo que las técnicas aplicadas en la etapa de extracción de conocimiento deben ser diferentes a las aplicadas con los datos personales y académicos de los alumnos. En un principio, se buscará la relevancia de cada uno de los tópicos y la posibilidad de descubrir agrupamientos subyacentes entre ellos. Estos resultados ayudarán a entender mejor el perfil de los alumnos de nuestra facultad.

Modelado de Perfiles de Alumnos y Sistemas de Recomendación

Uno de los puntos centrales de esta línea de investigación consiste en modelar perfiles dinámicos a través de información provista por los entornos de enseñanza y aprendizaje que usan las diferentes cátedras de la facultad, información de los préstamos de la biblioteca, actividad en las redes sociales, entre otros [7] [8] [9] [10].

Estos perfiles permitirán proporcionar información para un sistema de recomendación [11] [12] [13].

El sistema recomendador se desarrollará como una extensión de los entornos de aprendizaje, que utilizando la información de los perfiles obtenidos pueda recomendar materiales educativos [14].

Líneas de Investigación, Desarrollo e Innovación

- Estudio de técnicas de agrupamiento aplicables a datos masivos.
- Estudio de distintas técnicas de preprocesamiento aplicables a Minería de Textos.
- Estudio, análisis y comparación de diferentes técnicas de visualización para grandes volúmenes de datos.
- Revisión y análisis de técnicas específicas de Learning Analytics.
- Estudio y desarrollo sobre modelado de perfiles dinámicos de los alumnos según sus características más relevantes.
- Estudio de algoritmos de recomendación.
- Desarrollo de un sistema recomendador de materiales educativos utilizando la información de los perfiles obtenidos.

Resultados y Objetivos

- Detección de las características más relevantes de los estudiantes aportando información útil en relación a los datos personales y el rendimiento académico.
- Desarrollo de una prueba de concepto que arroja información preliminar adecuada para predecir la condición de no

regularidad de un estudiante y la incidencia de su situación laboral.

- Estudio y aplicación de técnicas de Minería de Datos para datos masivos con el objetivo de analizar información de las redes sociales.
- Estudio de las técnicas existentes para diseñar un sistema recomendador a partir de perfiles dinámicos de los alumnos.

Formación de Recursos Humanos

Dentro de los temas involucrados en esta línea de investigación, actualmente hay 2 investigadores realizando su doctorado, y uno de ellos a su vez, finalizando una maestría.

Referencias

[1] Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(6), (pp. 601-618).

[2] Díaz, Francisco Javier; Lanzarini, Laura Cristina; Charnelli, María Emilia; Baldino, Guillermo; Schiavoni, María Alejandra; Amadeo, Ana Paola. Analítica del aprendizaje y la personalización de la educación. *XVII Workshop de Investigadores en Ciencias de la Computación*. 2015, Salta.

[3] Charnelli, E. Lanzarini, L. Baldino, G. Diaz, F. Determining the profiles of young people from Buenos Aires with a tendency to pursue computer science studies. *XX Congreso Argentino de*

Ciencias de la Computación. La Matanza, 2014.

[4] Lanzarini, L.; Charnelli, M.E.; Diaz, J. Academic performance of university students and its relation with employment. *Latin American Computing Conference (CLEI), Perú 2015. IEEE Conference Publications*.

[5] Lanzarini, Laura Cristina; Charnelli, María Emilia; Baldino, Guillermo; Díaz, Javier F. Selección de atributos representativos del avance académico de los alumnos universitarios usando técnicas de visualización. Un caso de estudio. *Revista: TE & ET*; no. 15 . ISSN: 1850-9959. p. 42-50. 2015

[6] KAYA, Tugberk; BICEN, Huseyin. The effects of social media on students' behaviors; Facebook as a case study. *Computers in Human Behavior*, 2016, vol. 59, p. 374-379.

[7] W. Paireekreng and T. Prexawanprasut, "An integrated model for learning style classification in university students using data mining techniques," *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2015 12th International Conference on*, Hua Hin, 2015, pp. 1-5.

[8] Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Leony, D., & Kloos, C. D. (2015). ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform. *Computers in Human Behavior*, 47, 139-148.

- [9] Chrysafiadi, K., & Virvou, M. (2015). Student Modeling for Personalized Education: A Review of the Literature. In *Advances in Personalized Web-Based Education* (pp. 1-24). Springer International Publishing.
- [10] Mota, P., Melo, F., & Coheur, L. (2015, May). Modeling Students Self-Studies Behaviors. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (pp. 1521-1528). International Foundation for Autonomous Agents and Multiagent Systems.
- [11] Drachsler, H., Verbert, K., Santos, O. C., & Manouselis. Panorama of recommender systems to support learning. In *Recommender systems handbook* (pp. 421-451). Springer US, 2015.
- [12] Tejeda-Lorente, Á., Bernabé-Moreno, J., Porcel, C., Galindo-Moreno, P., & Herrera-Viedma, E. (2015). A Dynamic Recommender System as Reinforcement for Personalized Education by a Fuzzly Linguistic Web System. *Procedia Computer Science*, 55, 1143-1150.
- [13] LI, Yanyan, et al. Designing a Learning Recommender System by Incorporating Resource Association Analysis and Social Interaction Computing. *En State-of-the-Art and Future Directions of Smart Learning*. Springer Singapore, 2016. p. 137-143.
- [14] J. Díaz, A. Schiavoni, P. Amadeo, M. E. Charnelli, Búsqueda personalizada de recursos educativos abiertos basada en el perfil del alumno dentro de un entorno educativo, *XVI Workshop de Investigadores en Ciencias de la Computación*, Tierra del Fuego, 2014.

Minería de Textos y de la Web

Leticia Cagnina*, Edgardo Ferretti, M. Paula Villegas, M. José Garcarena,
Sergio Burdisso**, Darío Funez, Carlos Velázquez, Marcelo Errecalde

Laboratorio de Investigación y Desarrollo en Inteligencia Computacional

Departamento de Informática, Universidad Nacional de San Luis

Ejército de los Andes 950 - (D5700HHW) San Luis - Argentina

e-mails de contacto: {lcagnina, ferretti, merreca}@unsl.edu.ar

Resumen

Este artículo describe, brevemente, las tareas de investigación y desarrollo que se están llevando a cabo en la línea de investigación “Minería de Textos y de la Web” en el marco del proyecto “Aprendizaje automático y toma de decisiones en sistemas inteligentes para la Web”. La línea aborda diversas áreas vinculadas a la ingeniería del lenguaje natural, como por ejemplo el Procesamiento del Lenguaje Natural (PLN), la Lingüística Computacional, la Minería de Textos, la Minería de la Web y la recuperación de información de la Web. En el contexto de este proyecto por lo tanto, esta línea se centra en todos los problemas vinculados con el desarrollo de herramientas inteligentes para la extracción, análisis y validación de contenido Web, que incluyen: representación de documentos y usuarios de la Web, medidas de calidad de información para el contenido Web, técnicas abiertas de extracción de información para la Web, algoritmos de categorización supervisados, semi-supervisados y no supervisados y caracterización de usuarios, entre otros.

Palabras clave: Minería de Textos, Minería de la Web, Lingüística Computacional, Procesamiento del Lenguaje Natural

*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

**Becario de CONICET

Contexto

La línea de investigación “Minería de Textos y de la Web” es una de las tres líneas del proyecto titulado “Aprendizaje automático y toma de decisiones en sistemas inteligentes para la Web”, un nuevo proyecto que será presentado este año como continuación del Proyecto de Investigación Consolidado (PROICO) titulado “Herramientas y mecanismos para la toma de decisiones en agentes inteligentes artificiales”. Este último proyecto, aprobado por evaluadores externos a la UNSL, se desarrolla en el *Laboratorio de Investigación y Desarrollo en Inteligencia Computacional* (LIDIC) de la UNSL y ha sido financiado en forma directa por la UNSL (PROICO 30312) y en forma indirecta por: a) el Programa de Incentivos (22/F237), b) la Comisión Europea de Investigación e Innovación, a través del programa Marie Curie Actions: FP7 People 2010 IRSES, c) el CONICET, a través de un investigador asistente y becas: dos de Doctorado y dos de Post-Doctorado asignadas a integrantes del proyecto y d) el Consejo Nacional de Ciencia y Tecnología (CONACYT-México) y otros organismos científicos del gobierno mexicano, en los que distintos integrantes han participado en tres proyectos de investigación como colaboradores externos.

Este proyecto posee además dos líneas de investigación denominadas: “Aplicaciones” y “Toma de decisiones y aprendizaje automá-

tico”; la primera enfocada en el uso del aprendizaje automático en psicología, educación y el cuidado de la salud y la segunda dedicada al desarrollo de modelos formales y mecanismos para la toma de decisiones y aprendizaje en agentes artificiales inteligentes. Es claro en este contexto, que muchos problemas y aplicaciones intersectan los alcances de más de una de las líneas de este proyecto, lo cual involucra un trabajo integrado y coordinado permanente a los fines de optimizar los recursos disponibles para la obtención de los objetivos propuestos.

Introducción

En la actualidad, la cantidad de información disponible en la Web crece a un ritmo exponencial. Mucha de esta información está almacenada en forma de documentos de texto generados por diferentes usuarios los cuales poseen diversas características. Ejemplos de este tipo de material es el producido en las redes sociales como Facebook, Google+, sitios de microblogging como Twitter y las innumerables facilidades de chats disponibles hoy en día. La posibilidad de analizar toda la información disponible significa un reto muy importante para los investigadores de las Ciencias Sociales, razón por la que surge la necesidad de contar con herramientas automáticas que permitan acceder, organizar y almacenar el caudal de material con el que se cuenta. En este sentido, la *caracterización del autor* (en inglés, *author profiling*) es la tarea que tiene como principal objetivo el análisis de los textos de un autor con la finalidad de obtener tanta información como sea posible respecto de la/s persona/s que escribieron dichos textos. Información relacionada a la edad, género, personalidad, demografía, idioma natal y antecedentes culturales [7], son algunos ejemplos del tipo de información que se puede extraer considerando sólo los textos de una persona. El determinar correctamente el perfil de un autor es un problema que tiene una amplia gama de aplicaciones que podrían impactar en nues-

tras vidas de forma considerable. Por ejemplo, en marketing, el detectar características específicas de los usuarios (por ejemplo el género de personas que gustan de cierto producto), permitiría mostrar sólo ciertos tipos de productos a sólo ciertos tipos de usuarios. De manera similar, en el área de Business Intelligence, conocer qué tipo de personas son las interesadas en determinados servicios podría significar el éxito o fracaso de la empresa. Más interesante aún, en el campo forense, el reconocimiento del perfil lingüístico de acosadores (por ejemplo pedófilos) puede significar el hecho de identificar e incluso sentenciar a los sospechosos [20, 12].

Wikipedia, la enciclopedia en línea de libre acceso más popular e importante de todos los tiempos, es otra de las fuentes fundamentales de información en la Web. Sin embargo, su popularidad también conlleva un reto crucial: mejorar continua y sistemáticamente la calidad de los textos que la componen. Este aspecto no es casual si consideramos que los autores que contribuyen con Wikipedia son heterogéneos, en cuanto al nivel de educación, edad, cultura, habilidades del lenguaje y especialización en un área. De allí la importancia de poder identificar de manera automática ciertos aspectos de calidad como por ejemplo: la exactitud, fiabilidad y relevancia [21] de la información publicada. Diferentes herramientas han sido propuestas para la clasificación de los documentos de Wikipedia, como así también diferentes métricas para evaluar la presencia o no de diferentes fallas de calidad.

En la siguiente sección, se describen los principales enfoques desarrollados por los integrantes del grupo en lo que respecta al estudio de la caracterización del autor, calidad de información en Wikipedia y algunas extensiones de trabajos ya desarrollados como la clasificación no supervisada de textos cortos.

Desarrollo e Innovación

En términos generales, la línea “Minería de Textos y de la Web” se desarrolló siguiendo tres aristas bien marcadas. Cada una de ellas se detalla brevemente a continuación.

Mecanismos automáticos para la caracterización del autor (CA)

La CA basada exclusivamente en las características presentes en el texto que una persona ha escrito, ha sido una tarea muy interesante de llevar a cabo. Se han obtenido buenos resultados con técnicas estilográficas como los n -gramas de caracteres o algunas más avanzadas de segundo orden [23, 16] para la representación de los documentos. También el uso de *perfiles* con las características más importantes de cada grupo etario ha arrojado buenos resultados [15]. Actualmente, se trabaja en la búsqueda de nuevas estrategias de representación que consideren al *usuario* en un contexto más general que como el simple autor de un documento. La idea en este caso es considerar e integrar toda aquella información disponible que surge de la interacción del usuario con los medios sociales. No sólo consideraremos atributos léxicos, estilométricos o socio-lingüísticos presentes en los documentos, sino también atributos *multi-modales* como los derivados del grafo de contactos de un usuario, imágenes y videos que comparte en la Web, etc.

Calidad de la información en la web

Debido al fácil acceso a la información que existe en la actualidad a través de diferentes recursos, la evaluación de la calidad de la información en la Web se ha convertido en una tarea muy importante. Día a día tanto las personas comunes como empresas y entidades gubernamentales o privadas toman decisiones basándose en la información disponible en la Web. Esto, sumado al notable incremento de información disponible en Internet ha provocado una necesidad imperiosa de evaluar la calidad de dicha información de forma automática.

En este sentido se ha trabajado en la identificación y definición de diferentes aspectos relacionados a la calidad de información del contenido Web como confiabilidad, objetividad, especificidad, etc. Para ello se utilizaron como referencia diferentes propuestas existentes para el área de calidad de información en la Web [24, 18, 8]. Se desarrollaron características (*features*) basadas en información factual [17] y variantes del algoritmo PU-learning [19] que obtuvieron resultados muy interesantes en la clasificación de fallas de calidad en el contexto de Wikipedia [14, 13]. PU-learning es un algoritmo perteneciente al paradigma de aprendizaje semi-supervisado, ya que utiliza archivos no etiquetados para ayudar al clasificador en la distinción de la clase positiva. Los enfoques de clasificación one-class [22] también pertenecen a este paradigma de aprendizaje y en particular, la predicción de fallas de calidad en Wikipedia ha sido caracterizada como un problema one-class [1, 6, 4, 5, 2], por el grupo de investigación alemán¹ que dio origen a una línea de investigación que lleva el mismo nombre. De acuerdo con los reportes realizados por Anderka et al. [1, 2] existen diez fallas de calidad que comprenden aproximadamente el 75 % de documentos en Wikipedia, y es por eso que su predicción ha sido motivo de investigaciones recientes, principalmente a partir de la primera Competencia Internacional de Predicción de Fallas de Calidad en Wikipedia [3], realizada en el año 2012, en la que nuestro grupo obtuvo los mejores resultados.

Actualmente, estamos extendiendo medidas de calidad basadas en información factual, de manera tal de detectar fallas de calidad específicas. En este contexto, se están realizando pruebas con el subconjunto de fallas de calidad de Wikipedia en inglés denominado *Original Research* (una de las diez fallas más importantes, mencionadas precedentemente) para determinar la efectividad de este tipo de features.

¹<https://www.uni-weimar.de/en/media/chairs/webis/home/>

Categorización no supervisada

Con el objetivo de extender los trabajos previos de algunos de los integrantes del grupo en relación al clustering de textos cortos [10] a textos más generales, se establecieron dos líneas de trabajo. En primer lugar, se extendieron estos trabajos a documentos de longitud arbitraria [11]. Luego, se buscó determinar el grado de escalabilidad de los métodos ya desarrollados, y en particular aquellos basados en enfoques de Inteligencia Colectiva [9]. En este mismo contexto, estamos analizando implementaciones más eficientes de algoritmos como *Sil-Att* [11], mediante modificaciones de la implementación del Coeficiente de Silueta y una versión adaptativa de este mismo coeficiente.

Formación de Recursos Humanos

Trabajos de tesis vinculados con las temáticas descritas previamente:

- 2 tesis de Licenciatura defendidas en 2015.
- 2 tesis de Licenciatura a iniciarse en Marzo de 2016.
- 2 tesis de Maestría en ejecución.
- 1 tesis de Doctorado en ejecución con una beca de CONICET.

Referencias

- [1] M. Anderka. *Analyzing and Predicting Quality Flaws in User-generated Content: The Case of Wikipedia*. Dissertation, Bauhaus-Universität Weimar, 2013.
- [2] M. Anderka and B. Stein. A Breakdown of Quality Flaws in Wikipedia. In C. Castillo, Z. Gyongyi, A. Jatowt, and K. Tanaka, editors, *2nd Joint WICOW/AIRWeb Workshop on Web Quality*, pages 11–18. ACM, 2012.
- [3] M. Anderka and B. Stein. Overview of the 1st International Competition on Quality Flaw Prediction in Wikipedia. In P. Forner, J. Karlgren, and C. Womser-Hacker, editors, *Working Notes Papers of the CLEF 2012 Evaluation Labs*, 2012.
- [4] M. Anderka, B. Stein, and N. Lipka. Detection of Text Quality Flaws as a One-class Classification Problem. In B. Berendt, A. de Vries, W. Fan, C. Macdonald, I. Ounis, and I. Ruthven, editors, *20th ACM International Conference on Information and Knowledge Management*, pages 2313–2316. ACM, 2011.
- [5] M. Anderka, B. Stein, and N. Lipka. Towards Automatic Quality Assurance in Wikipedia. In S. Srinivasan, K. Ramamritham, A. Kumar, M. Ravindra, E. Bertino, and R. Kumar, editors, *20th International Conference on World Wide Web*, pages 5–6. ACM, 2011.
- [6] M. Anderka, B. Stein, and N. Lipka. Predicting Quality Flaws in User-generated Content: The Case of Wikipedia. In B. Hersh, J. Callan, Y. Maarek, and M. Sanderson, editors, *35th International ACM Conference on Research and Development in Information Retrieval*, pages 981–990. ACM, 2012.
- [7] S. Argamon, M. Koppel, J. W. Pennebaker, and J. Schler. Automatically profiling the author of an anonymous text. *Commun. ACM*, 52(2):119–123, 2009.
- [8] R. Baeza-Yates. User generated content: how good is it? In *3rd Workshop on information credibility on the Web*, 2009.
- [9] L. C. Cagnina, M. Errecalde, D. Ingaramo, and P. Rosso. An efficient particle swarm optimization approach to cluster short texts. *Information Science*, 265:36–49, 2014.

- [10] M. Errecalde, D. Ingaramo, and P. Rosso. ITSA*: An effective iterative method for short-text clustering tasks. In *Proc. of the 23rd International Conference on Industrial Engineering and other Applications of Applied Intelligent Systems, IEA/AIE 2010*, volume 6096 of *LNCS*, pages 550–559. Springer-Verlag, 2010.
- [11] M. L. Errecalde, L. C. Cagnina, and P. Rosso. Silhouette + attraction: A simple and effective method for text clustering. *Natural Language Engineering*, FirstView:1–40, 2 2016.
- [12] H. J. Escalante, E. Villatoro-Tello, A. Juarez, M. M. y Gomez, and L. Vil-lasenor. Sexual predator detection in chats with chained classifiers. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 46–54. ACL, 2013.
- [13] E. Ferretti, M. L. Errecalde, M. Anderka, and B. Stein. On the use of reliable-negatives selection strategies in the pu learning approach for quality flaws prediction in wikipedia. In H. Decker, L. Lhotská, S. Link, M. Spies, and R. R. Wagner, editors, *DEXA Workshops*, pages 211–215. Springer, 2014.
- [14] E. Ferretti, D. H. Fusilier, R. Guzmán-Cabrera, M. M. y Gómez, M. Errecalde, and P. Rosso. On the use of pu learning for quality flaw prediction in wikipedia. In P. Forner, J. Karlgren, and C. Womser-Hacker, editors, *CLEF (Online Working Notes/Labs/Workshop)*, 2012.
- [15] D. G. Funez, L. Cagnina, and M. L. Errecalde. Determinación de género y edad en blogs en español mediante enfoques basados en perfil. In *XVIII Congreso Argentino de Ciencias de la Computación*, pages 984–993, 2013.
- [16] M. J. Garciarena Ucelay, M. P. Villegas, L. C. Cagnina, and M. L. Errecalde. Cross domain author profiling task in spanish language: An experimental study. *Journal of Computer Science and Technology*, 41(2):122–128, 2015.
- [17] E. Lex, M. Voelske, M. Errecalde, E. Ferretti, L. Cagnina, C. Horn, B. Stein, and M. Granitzer. Measuring the quality of web content using factual information. In *Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality*. ACM, 2012.
- [18] A. Lih. Wikipedia as participatory journalism: reliable sources? metrics for evaluating collaborative media as a news resource. In *5th international symposium on online journalism*, 2004.
- [19] B. Liu, Y. Dai, X. Li, W. Lee, and P. Yu. Building text classifiers using positive and unlabeled examples. In *3rd IEEE international conference on data mining*. IEEE Computer Society, 2003.
- [20] I. McGhee, J. Bayzick, A. Kostostathis, L. Edwards, A. McBride, and E. Jakubowski. Learning to identify internet sexual predation. *International Journal of Electronic Commerce*, 15(3):103–122, 2011.
- [21] T. Redman. *Data Quality for the Information Age*. Artech House, 1996.
- [22] D. Tax. *One-class classification*. Ph.d. thesis, Delft University of Technology, 2001.
- [23] M. P. Villegas, M. J. Garciarena Ucelay, M. L. Errecalde, and L. Cagnina. A spanish text corpus for the author profiling task. In *XX Congreso Argentino de Ciencias de la Computación*, pages 621–630, 2014.
- [24] R. Y. Wang and D. M. Strong. Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 1996.

FRAMEWORK PARA DATA MINING EDUCATIVO: FORMALIZACION Y APLICACIONES

Marcelo Omar Sosa¹, Carlos Iván Chesñevar², Eugenia Cecilia Sosa Bruchmann¹

¹Departamento Computación/Facultad de Ciencias Exactas y Naturales/Universidad Nacional de Catamarca
Av. Belgrano N° 300 - Planta alta - C.P: 4700 - San Fernando del Valle de Catamarca
Teléfono: 0383- 4425610 /4420900

²Instituto de Ciencias e Ingeniería de la Computación (Conicet –U.N.S.) / Facultad de Ciencias e Ingeniería
Universidad Nacional del Sur
Av. Alem 1253 -B8000CPB Bahía Blanca, Argentina
Teléfono: 0291- 4595135 ext. 2610

sosamod1@hotmail.com, cic@cs.uns.edu.ar, sosab_ec@hotmail.com

Resumen

Las técnicas de data mining permiten analizar grandes volúmenes de datos en búsqueda de información oculta y relevante para la toma de decisiones. Estas se aplican en diversos campos en donde se almacenan grandes volúmenes de datos de las actividades realizadas y cuyo procesamiento no puede realizarse utilizando otras técnicas. En el caso de datos obtenidos de procesos educativos, éstos presentan características particulares que requieren técnicas y formas de interpretación de resultados especiales por lo que dio origen a la rama de data mining denominada Educational data mining o E.D.M. por sus siglas en inglés. El proceso educativo moderno incorpora la tecnología como medio de comunicación y de desarrollo de actividades fuera del ámbito del aula, si bien existen diferentes tipos, el más utilizado es el denominado blended learning ya que representa una adecuada combinación de actividades virtuales y presenciales con el objetivo de enriquecer el proceso. La actividad educativa así desarrollada genera grandes volúmenes de datos, su procesamiento con técnicas de data mining y la interpretación de los resultados obtenidos requiere la creación de un framework que agrupe las técnicas,

prácticas y criterios que sean más adecuadas para el procesamiento de este tipo especial de datos y que ayuden al docente a su aplicación e interpretación.

Palabras clave: Técnicas de data mining, Educational data mining.

Contexto

El presente trabajo de investigación se desarrolla dentro de las líneas prioritarias del centro de investigación en estadística aplicada (CEA-Fa.C.E.N.-U.N.Ca.) que fuera creado en el año 2015 en la Universidad Nacional de Catamarca Facultad de Ciencias Exactas y Naturales en el marco del Sistema de Investigación, desarrollo e innovación. Además los docentes están a cargo de asignaturas de las carreras del área informática de la Facultad de Ciencias Exactas y Naturales (Fa.C.E.N.) de la Universidad Nacional de Catamarca (U.N.Ca.). Estas asignaturas se dictan en el campus central en la Capital de la provincia de Catamarca como así también en las subsedes del interior (Departamento Ancasti) y en la provincia de Tucumán (San Miguel de Tucumán).

Introducción

El data mining está compuesto por numerosas técnicas que tienen como objetivo el buscar patrones e información relevante y oculta dentro de los datos almacenados en grandes bases de datos [6]. Actualmente con los avances en tecnología se han desarrollado variadas aplicaciones que permiten y facilitan la aplicación de técnicas de data mining y el procesamiento de grandes volúmenes de datos que son la materia prima para esta disciplina. La aplicación de estas técnicas como así también la utilización de las herramientas disponibles requiere que el usuario cuente con un cierto grado de conocimiento sobre el tema para la correcta selección y aplicación de las mismas. En los últimos años se ha venido desarrollando el área de aplicación de estas técnicas en datos obtenidos de procesos educativos con la utilización de diferentes medios tecnológicos, entre los que se encuentra la mediación con internet como una forma de facilitar el acceso y aumentar la carga horaria de las asignaturas. Los numerosos accesos para la realización de actividades, las relaciones que se crean y desarrollan, como las interacciones entre alumnos y docentes generan este gran volumen de datos de los cuales pueden extraerse información que correctamente interpretada puede favorecer al mejoramiento del proceso educativo guiando al docente en la toma de decisiones. Esta rama del data mining que procesa datos educativos se denomina Educational data mining (EDM) [1], y tiene como objetivo el descubrimiento de información a través de la aplicación de las técnicas a los datos relacionados con el desempeño pedagógico generados en plataformas LMS y también los datos de las actividades presenciales realizadas por los alumnos. Se desea analizar estos datos para extraer información por diversas razones como lo son:

- comprender la forma en que aprenden los alumnos.
- descubrir la mejor forma de organizar los materiales y actividades de la asignatura,

- determinar cuáles atributos son más representativos para ser utilizados en las predicciones.
- desarrollar nuevas tipologías de estudiantes y ajustar las existentes.
- determinar la formación de grupos.
- descubrir patrones de comportamientos.
- modificar las estrategias pedagógicas. entre varias otras.

La aplicación de las técnicas de data mining permite superar las dificultades que presentan otras metodologías como las estadísticas en cuanto al manejo de grandes volúmenes de datos como así también al elevado número de variables que deben analizarse durante el procesamiento. El análisis de datos educativos puede realizarse desde distintos enfoques de acuerdo a las técnicas que se apliquen, estas son:

Descripción: Este grupo presenta como principal objetivo el de describir las características más representativas de los datos en busca de un modelo que englobe los diferentes tipos analizados. Cuando estas técnicas se aplican a datos educativos lo que hace es caracterizar a los alumnos según su desempeño académico en las actividades propuestas por el docente relacionadas con el contenido específico de la asignatura [3].

Predicción: Las técnicas que se agrupan es este tipo buscan establecer un modelo que represente a los datos y que permita estimar los valores que pueden tomar a futuro las variables analizadas, en el caso de su aplicación a datos educativos los atributos iniciales de los alumnos son analizados por estas técnicas obteniéndose resultados que le predicen al docente como puede llegar a ser el desempeño pedagógico durante las actividades a realizar. Estas predicciones se basan en el análisis de los datos iniciales y resultados intermedios que van obteniendo los alumnos durante el desarrollo del contenido de la asignatura [4].

Segmentación: Se basan en encontrar conjuntos de datos con características

similares, conformando grupos de comportamientos similares y que sean diferentes entre sí. Cuando se procesan datos educativos los grupos se encuentran separados basándose en el comportamiento de los alumnos que presentan durante el desarrollo de las actividades de la asignatura [5].

Cada grupo de técnicas se aplica y representa un método o un enfoque conceptual para extraer la información de los datos. Estas pueden implementarse por medio de varios algoritmos que indican los pasos a seguir para la aplicación de cada técnica. Gracias a ellas pueden predecirse resultados generando modelos predictivos o encontrar relaciones que generan modelos descriptivos. Los resultados del procesamiento de los datos educativos con data mining tienen como destinatario al docente, que además, desde el punto de vista de la conveniencia representa el mejor usuario para realizar el procesamiento por participar del proceso y ser quien conoce en profundidad el campo de la disciplina. Como podemos observar tanto las técnicas como su aplicación tienen en general un grado de dificultad que requiere conocimientos por lo menos básicos sobre data mining, por ello en ayuda al docente, se ve la necesidad de desarrollar un framework que oriente y lo ayude en esta tarea. El desarrollo de dicho framework permitirá guiar a los docentes interesados en aplicar las técnicas de data mining durante el procesamiento de los datos, sugiriendo las técnicas más adecuadas, estableciendo prácticas y criterios para la interpretación de los resultados. Se espera que el aporte del data mining educativo al proceso, mejore en gran medida su efectividad. Permitiendo al docente autoregular los contenidos, las prácticas pedagógicas como así también poder acercarle herramientas que permitan una clasificación a priori del perfil dominante de los alumnos como así también presentar los patrones emergentes de las actividades desarrolladas durante el cursado de la asignatura.

Metodología

Se pretende con esta investigación el generar un marco de referencia para la consulta de docentes interesados en innovar en el proceso educativo. El estudio de los resultados que presentan de la aplicación de las diferentes técnicas aplicándolas a datos educativos serán valorados desde el punto de vista del docente. Particularmente para el estudio se utilizarán datos de asignaturas con contenidos de programación de las diferentes carreras y subsedes de la facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Catamarca. Como finalización de esta investigación se estima el poder desarrollar un framework que pueda ser aplicado por docentes para su validación.

Conclusión

Se espera establecer un conjunto de conceptos, prácticas y criterios que permitan procesar e interpretar la información obtenida de la aplicación de técnicas de data mining a datos obtenidos del proceso educativo en asignaturas relacionadas con contenidos de programación. Con el objetivo principal mejorar el proceso educativo mediante una adecuada modificación de metodologías y técnicas relacionadas con las características específicas del perfil dominante de los alumnos cursantes actuales.

Líneas de investigación y desarrollo

El presente trabajo se enmarca dentro de la investigación realizada para el desarrollo de la tesis de Maestría en Ciencias de la Computación, en donde se investigan los aportes del data mining al proceso educativo. Con el objetivo de mejorar dicho proceso, se presentan nuevos enfoques para el procesamiento de datos educativos utilizando técnicas y herramientas de data mining. Particularmente se trabaja con datos obtenidos del desarrollo de las actividades de asignaturas con contenido de programación por lo que presentan características distintivas de datos de otras asignaturas de la misma carrera.

El tema que se presenta viene profundizándose mediante el estudio continuo y con la

presentación de diferentes trabajos en reuniones científicas en donde se muestran los avances realizados y los posibles resultados que se esperan de la investigación. La propuesta se enmarca dentro de los temas de investigación del Centro de Estudios Estadísticos creado en la Universidad Nacional de Catamarca, Facultad de Ciencias Exactas y Naturales (Catamarca, Argentina).

Resultados y Objetivos

Los resultados esperados para esta investigación son los de establecer un conjunto de conceptos, prácticas y criterios para el procesamiento e interpretación de los resultados del procesamiento de datos educativos con técnicas y herramientas de data mining.

Tiene como objetivos principales:

- Conformar un conjunto de técnicas de data mining que permitan mediante su aplicación la obtención de información para la toma de decisiones por parte del docente.
- Procesar los datos educativos con la finalidad de comprender la forma en que aprenden y realizan las actividades los alumnos.
- Obtener sugerencias para mejorar el proceso educativo en relación con el aprendizaje de la programación.
- Establecer los criterios de selección de datos que aportan más información en el procesamiento para establecer el perfil dominante de los alumnos en cada cohorte.
- Establecer criterios para la comprensión de la información que proporcionan las técnicas de data mining por tratarse de un caso especial como lo son los datos educativos.
- Determinar el método de selección de atributos que sean más representativos de las características de los alumnos que cursan asignaturas con contenidos de programación.

Formación de Recursos Humanos

Los autores del trabajo se encuentran en la etapa de desarrollo de sus tesis de posgrado en carreras relacionadas con el tema de investigación, como la Maestría en Ciencias de la Computación en donde el Mgter. Marcelo Sosa desarrolla actualmente la tesis denominada: ***“Aportes de data mining a la mejora del proceso educativo con blended learning: formalización y experimentaciones”*** bajo la dirección del Dr. Carlos Chesñevar perteneciente a la Universidad Nacional del Sur (U.N.S). Además el investigador se encuentra en la etapa de planificación de su tesis doctoral en el área de minería de datos en el Doctorado en Ciencias dictado en la Facultad de Ciencias Exactas y Naturales (Fa.C.E.N.) en convenio con la Universidad Nacional del Sur (U.N.S). La Docente Investigadora Lic. Eugenia Sosa Bruchmann desarrolla su tesis en la carrera Especialización en Ingeniería en Software de la Universidad Nacional de San Luis denominada ***“La experiencia del usuario desde un nuevo enfoque para el desarrollo de productos interactivos: el comportamiento emocional del usuario y la importancia de los atributos estéticos”*** dirigida por el Dr. Germán Montejano. Los docentes investigadores desarrollan actividades de dirección de tesis de la carrera de Licenciatura en Tecnología Educativa de los siguientes alumnos: Varela Marino del Valle cuya tesis se denomina ***“Análisis del impacto de un aula virtual en el proceso de enseñanza y aprendizaje en la Escuela de Educación Técnica N° 7 “José Alsina Alcobert”***, y Martín Fabián Molina cuya tesis se denomina: ***“Estudio de la implementación del uso de TIC’s en la enseñanza de la educación vial en el sistema educativo municipal de San Fernando del Valle de Catamarca”***.

Además desarrollan las siguientes actividades:

- Dirección de proyectos de investigación de voluntariado y pertenecientes a la facultad a la cual pertenecen.

- Integración de equipo de investigación de centro de investigación en Estadística de la Facultad de Ciencias Exactas y Naturales del U.N.Ca.
- Producción de artículos científicos para su presentación en congresos locales, nacionales e internacionales.
- Participación de los integrantes en cursos de actualización y posgrado en el área de estudio.
- Integrantes de la revista de ciencias de la Facultad de Ciencias Agrarias de la U.N.Ca.
- La actualización y capacitación permanente de los investigadores en talleres o workshop relacionadas con el tema del trabajo.
- La participación de los investigadores como consultores en proyecto afines que se desarrollan en la Facultad de Ciencias exactas y Naturales en distintas áreas.
- Examinadores de trabajos finales en las diferentes carreras que se dictan en la Fa.C.E.N. de las U.N.Ca.
- Dirección de tesis y tesinas finales de las carreras de computación, informática y Licenciatura en tecnología educativa.
- La planificación de seminarios para docentes en temas relacionados con la investigación y resultados obtenidos en la investigación.
- Participación en convenios con la Facultad de Tecnología para el desarrollo de estudios del área de datamining.

REFERENCIAS

1. C. Romero and S. Ventura, "Educational data mining: A Survey From 1995 to 2005", *Expert System with Applications*, vol. 33, pp. 135-146, 2007.
2. Lavrac, N., Kavsec, B., Flach, P. and Todorovski, L., "Subgroup discovery with CN2-SD". *Journal of machine learning research*. 2004.
3. Jain A.K. and Dubes R.C. "Algorithm for clustering data. 1998. Englewood Cliffs. N.J. Prentice Hall.
4. Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., "The KDD process for extracting useful knowledge from volumes of data". *Communication of ADM* 1996.
5. Agrawal, R., Imielinski, T. and Swami, A.N., "Mining Association Rules between set of item in large databases". In *International conference on management of data*. 1993. Washington D.C. ACM Press.
6. Solarte Martinez, Guillermo Roberto, Ocampo S., Carlos Alberto. *Técnicas de clasificación y análisis de representación del conocimiento para problemas de diagnósticos*. *Cientia Et Technica [En Linea]* 2009, Xv (Agosto-Sin Mes) : [Fecha De Consulta: 2 de marzo de 2016].
7. Gabriel Páramo and Carlos Correa, "Deserción Estudiantil Universitaria. Conceptualización," *Medellín, Revista Abril - Mayo – Junio 1999*.
8. Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, "From Data Mining to Knowledge Discovery in Databases," *AI Magazine*, pp. 37-54, 1991.
9. Daniel T Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. New York: John Wiley & Sons, 2004.
10. Brijesh Kumar Baradwaj and Saurabh Pal, "Mining Educational Data to Analyze Students' Performance," in *International Journal of Advanced Computer Science and Applications*, India, 2011, pp. Vol. 2, No. 6.
11. Baker Ryan and Kalina Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," *JEDM - Journal of Educational Data Mining*, vol. 1, no. 1, Octubre 2009.
12. Jonathan E. Freyberger, Neil Heernan, and Carolina Ruiz. *Using association rules to guide a search for best fitting transfer models of student learning*. Master's thesis, Worcester Polytechnic Institute, 2004.
13. Merceron and K. Yacef. *Educational data mining: a case study*. *Process of 12th. Conference on Artificial Intelligence in Education (AIED03)*, page 467 .2005.
14. Maria Delia Grossi. "Reglas de predicción aplicables al diseño de un curso de computación". Marzo 2008.
15. Erwin Sergio Fischer Angulo. "Modelo para la automatización del proceso de determinación de riesgo de deserción en estudiantes universitarios". Santiago de Chile 2012.
16. Pedro Gonzalez Garcia. "Aprendizaje evolutivo de reglas difusas para la descripción de subgrupos". Granada España. 2007.
17. Cristoban Romero, Sebastian Ventura, Nikola Pechenizkiy and Rayan Beker. "Handbook of educational data mining". Chapman & Hall CRC press. 2011

FRAMEWORK BASADO EN ONTOLOGÍAS PARA LA RECUPERACIÓN DE IMÁGENES MÉDICAS

Matías Agüero¹, Carlos Alvez², Marcela Vegetti¹

¹Facultad Regional Santa Fe – Universidad Tecnológica Nacional
Lavaise 610, Santa Fe, Santa Fe, S3002GJC, Argentina

²Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos
Av. Tavella 1424, Concordia, Entre Ríos - CP 3200
mjaguero@gmail.com, caralv@fcad.uner.edu.ar, mvegetti@santafe-conicet.gov.ar

Resumen

Las imágenes digitales en medicina son una base fundamental para poder brindar un diagnóstico acertado sobre el estado de salud de los pacientes. Su adquisición, almacenamiento, traslado y visualización son un desafío por la gran cantidad de éstas que existen en los centros de salud. Por esto es importante la definición de modelos de almacenamiento en bases de datos, tanto de las imágenes como de sus metadatos para una adecuada gestión de las mismas, principalmente en lo que refiere a recuperación e interoperabilidad. El estándar DICOM establece un mismo formato de imagen digital para todos los estudios en medicina, y a la vez, incluye información del paciente, del estudio realizado, equipos y mucha otra información relacionada con la imagen. La finalidad de este trabajo es la utilización de Ontologías para relacionar la información de todas las imágenes y recuperar las similares a través del uso de consultas semánticas sobre las Ontologías creadas con las imágenes DICOM.

Palabras clave: Base de datos OR, DICOM, Ontologías.

Contexto

Este trabajo consiste en el desarrollo de una tesis de maestría, que da continuidad al trabajo realizado en la tesis doctoral “Modelos para la recuperación de imágenes por similitud en Bases de Datos Objeto-Relacionales” [1], cuyo objetivo fue la recuperación de imágenes por similitud, sea por contenido físico como semántico, basado principalmente en modelos de representación de imágenes en Bases de Datos Objeto-Relacionales (BDOR) [2].

Introducción

Las imágenes digitales en medicina están reemplazando cada día más a los antiguos formatos físicos que ocupaban inmensos espacios destinados a almacenar placas, hojas de papel, y otros medios. Una vez almacenadas en depósitos no se volvían a consultar, ocasionando gastos y ocupando lugar en las instituciones médicas. Las imágenes digitales proporcionan diferentes ventajas sobre los antiguos formatos físicos desde el momento de la adquisición hasta el de su visualización. Entre ellas:

- Utilización de estándares en el formato de las imágenes.
- Reducción de espacios para el almacenamiento.
- Mejores tiempos de respuesta para la

disponibilidad de las mismas debido a una mejor organización.

- Posibilita llevar una mejor historia clínica de los pacientes.
- Las mismas pueden ser visualizadas, por diferentes especialistas, al mismo tiempo, en un lugar diferente al de donde se realizó el estudio, gracias a la posibilidad de comunicación que proporcionan.
- El paciente no necesariamente tiene que cargar con los estudios que se realizó.

Una imagen médica no tiene sentido sola, sino que es importante la información del contexto en el que se ha tomado la misma. En este sentido, el estándar DICOM (Tratamiento Digital de Imágenes y Comunicaciones en Medicina) [3], es el estándar que describe detalladamente los medios para dar formato e intercambiar imágenes e información entre diferentes dispositivos o equipos médicos de diagnóstico por imágenes. Además, en el contexto de una imagen DICOM podemos encontrarnos con datos del paciente, del estudio que se le está realizando, del equipo que ha efectuado el estudio, de las imágenes tomadas (número de tomas realizadas, separación entre cada imagen, dimensión de las imágenes, etc.). Asimismo, distintos estudios médicos e imágenes de un paciente están relacionados entre sí. Un estudio que se realiza a un paciente puede haber sido derivado de uno anteriormente realizado. Lo más importante es poder identificar a quien corresponde cada examen, el momento en el que se lo realizó y quien lo solicitó. Todo esto es posible a través del estándar DICOM, el cual propone organizar las imágenes de un paciente.

Actualmente existen sistemas que realizan la gestión de imágenes, conocidos como PACS (Sistemas de Almacenamiento y Comunicación de

Imágenes), en donde la comunicación en ambiente de red es la parte medular para el diseño de aplicaciones y se apoya en el protocolo DICOM para la gestión de la imagen diagnóstica. En los últimos años se ha producido una generalización a nivel mundial del uso de este protocolo.

Una herramienta útil para los profesionales en medicina al momento de efectuar el diagnóstico de un paciente, es lograr la recuperación de imágenes similares a la imagen obtenida en un estudio efectuado a dicho paciente. Para lo cual, además de contar con las imágenes en formato DICOM, también es necesario establecer relaciones entre las mismas de manera que, a partir de ciertas reglas o datos, poder recuperar otras imágenes y diagnósticos similares que se encuentren almacenados en la base de datos. La capacidad de las ontologías [4] de inferir nuevo conocimiento a partir de información explícita, hacen de las mismas una ayuda importante para la realización de estas búsquedas por similitud.

Una ontología describe un conjunto de entidades y sus relaciones, en donde se especifica formalmente las entidades en un dominio y sus propiedades (atributos). Permite que las personas y los sistemas de información accedan al conocimiento almacenado en su interior. Grandes y sofisticadas ontologías se han desarrollado para la genética y la anatomía humana. Actualmente hay cerca de 200 ontologías diferentes para una variedad de disciplinas biomédicas (<http://biportal.bioontology.org>). Las ontologías son "computables", es decir la información altamente estructurada que contienen puede ser utilizada para mejorar la recuperación de información, el descubrimiento de conocimiento y razonamiento automatizado.

Líneas de Investigación, Desarrollo e Innovación

La principal línea de investigación que se está trabajando es la gestión de imágenes en bases de datos OR (Objeto Relacionales). En este trabajo se trata la creación de modelos para la representación de datos relacionados con imágenes médicas, de manera que los mismos puedan consultarse, recuperarse y compararse de manera simple y eficiente. El uso de la tecnología OR se debe a que los modelos para la recuperación de imágenes, tienen estructuras de datos complejas y pueden ocasionar algunos inconvenientes si se trabaja con el modelo de datos relacional. En primer lugar, los metadatos relacionados a las imágenes son estructuras no atómicas, por lo que en las bases de datos relacionales (para respetar la primera forma normal), se deben separar en varias tablas, lo que hace menos eficiente el tratamiento.

El modelo OR [2] permite dar soluciones a las limitaciones de las bases de datos relacionales. En primer lugar, permite definir tipos de datos que pueden contener estructuras complejas como colecciones, objetos grandes, etc., sin la limitación de la primera forma normal. En segundo lugar, permite definir e implementar el comportamiento de los datos creando los métodos para la gestión de las instancias de los tipos definidos, y así facilitar el acceso seguro desde las aplicaciones que los utilicen. Por último, los sistemas de bases de datos OR, proveen paquetes para la gestión de imágenes DICOM [5] [6] y el tratamiento de datos semánticos [7].

Las facilidades mencionadas, posibilitan la creación de infraestructuras que extienden los servicios del sistema de gestión de base de datos [8].

Otra línea de investigación que abarca este trabajo, está relacionada con las ontologías. Una ontología es “una especificación formal y explícita de una conceptualización compartida” [9]. Una conceptualización es un modelo abstracto de algún fenómeno de la realidad que identifica los conceptos más relevantes de dicho fenómeno. Estos conceptos, así como las restricciones que pudieran existir sobre su utilización y sus relaciones deben estar explícitamente definidos en un lenguaje entendible por máquinas. Asimismo, la noción de compartido que menciona la definición se refiere a que una ontología captura conocimiento que no es privado a un único individuo sino consensuado por un grupo en un dominio determinado [10].

Desde las últimas décadas las ontologías se han vuelto relevantes y están siendo ampliamente utilizadas en diversos campos como la Ingeniería del Conocimiento, la Inteligencia Artificial y las Ciencias de la Computación, en aplicaciones relacionadas con la gestión del conocimiento, el procesamiento del lenguaje natural, el comercio electrónico, la información de integración inteligente, recuperación de información, diseño e integración de bases de datos, la bioinformática, la educación y la web Semántica. El crecimiento que se observa en el uso de las ontologías se debe en gran medida en la capacidad que brindan para la definición de la semántica de los términos en un dominio y para la inferencia de nuevo conocimiento basado en la conceptualización explícita y formalmente definida.

Las ontologías, entonces, permitirán enriquecer semánticamente los metadatos DICOM para poder inferir nuevo conocimiento sobre estas imágenes y lograr hacer más eficiente la búsqueda de imágenes por similitud de contenido semántico. Para lograr esto, esta línea de

investigación abordará un conjunto de actividades que van desde la búsqueda de ontologías, el anotado semántico de las imágenes hasta la especificación de las consultas necesarias para la búsqueda por similitud.

El relevamiento, análisis y selección de ontologías que permitan el anotado semántico de las imágenes es una meta central de este trabajo. En este sentido, se deberá enfocar la búsqueda en ontologías que permitan representar el estándar DICOM y en ontologías que permitan extender los metadatos DICOM con información respecto a enfermedades y diagnósticos. Se pretende focalizar esta búsqueda en ontologías que representen estándares, tesauros y catálogos ampliamente aceptados en el ámbito biomédico.

Una vez seleccionada las ontologías se deberá definir la manera en que se vincularan las mismas. Asimismo se deberá especificar la manera en que las imágenes DICOM almacenadas serán semánticamente anotadas con los conceptos de las ontologías. Finalmente, se deberán escribir las consultas semánticas que permitirán las búsquedas por similitud.

Resultados y Objetivos

El objetivo general de esta investigación es crear una arquitectura genérica que permita la recuperación de imágenes en medicina (formato DICOM) por similitud semántica, a través de la utilización de Ontologías. Los objetivos específicos se han alcanzado de manera parcial como se indica a continuación.

Objetivos alcanzados

- *Crear modelos en una base de datos Objeto-Relacional que soporten el estándar DICOM.* Se diseñaron y construyeron tablas, objetos y ontologías de la base de datos que

serán los encargados de almacenar y permitir la recuperación de las imágenes médicas por contenido semántico.

- *Crear ontologías en la base de datos para la búsqueda semántica de imágenes y diagnósticos.* Durante el proceso de investigación y diseño de una ontología para este trabajo se basó principalmente en la línea definida por el estándar DICOM, define como principales entidades a *Estudio, Serie e Imagen* (Fig. 1). Además se utilizaron ontologías ya existentes desarrolladas por otros investigadores y organizaciones, y de esta manera poder tomar ventajas de las mismas e integrarlas a lo desarrollado en esta investigación. Las ontologías utilizadas como complemento de lo desarrollado en este trabajo son las ontologías de Síntomas y Enfermedades Humanas. Las mismas las podemos encontrar en los sitios Berkeley.org¹ y BioPortal².

Objetivos pendientes

- *Crear un software para la inserción y recuperación de imágenes DICOM.* Este objetivo se logro de manera parcial. Hasta aquí se han implementado los modelos definidos en un sistema de gestión de base de datos objeto-relacional. Queda pendiente, finalizar con el desarrollo de una aplicación que permitirá almacenar imágenes DICOM y sus metadatos en el modelo semántico en la base de datos, y así, poder recuperar tanto imágenes como información con contenido semántico. Esta aplicación, permitirá visualizar las imágenes

¹ Berkeley Bioinformatics Open-source Projects
<http://www.berkeleybop.org/>

² Repositorio de ontologías bio-médicas:
<http://bioportal.bioontology.org/>

almacenadas en la base de datos, especificar los criterios de búsquedas y mostrar los resultados.

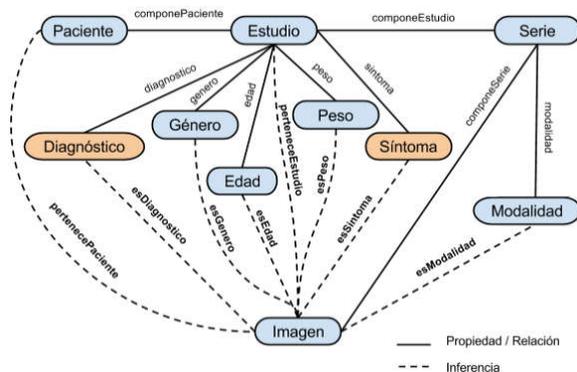


Fig. 1: Ontología en base al estándar DICOM.

Formación de Recursos Humanos

Este trabajo se encuentra enmarcado en el desarrollo de una tesis de maestría, con lo cual este plan de trabajo contribuirá a la formación del tesista en diferentes dominios: el tratamiento de imágenes y los estándares que describen el formato para el intercambio de las mismas, así como también, diversos aspectos en el campo de las ontologías y tecnologías semánticas. El equipo de trabajo está compuesto por el maestrando, junto con el Director y Co-Director de la misma. El tesista se encuentra realizando el proyecto final o tesis para la obtención de la Maestría en Ingeniería de Sistemas, la cual ya cursó en la Universidad Tecnológica Nacional (UTN). El estado de la tesis se encuentra en un alto porcentaje de avance, y los resultados finales, serán presentados en congresos nacionales o internacionales relacionados con el área Sistemas de Información, así como también, en revistas internacionales con referato. Estos trabajos servirán para divulgar los conocimientos obtenidos durante el trabajo de investigación.

Referencias

[1] Alvez Carlos E. “Modelos para la

recuperación de imágenes por similitud en Bases de Datos Objeto-Relacionales”. Tesis Doctoral. Santa Fe, Argentina, 2012. ISBN 978-987-33-2249-5.

- [2] Melton Jim, “(ISO-ANSI Working Draft) Foundation (SQL /Foundation)”, ISO/IEC 9075-2:2003 (E), United States of America (ANSI), 2003.
- [3] Página oficial de NEMA. <http://medical.nema.org/>
- [4] Mabotuwana, T., Lee, M.C., Cohen-Solal, E.V. (2013). An ontology-based similarity measure for biomedical data - Application to radiology reports. *Journal of Biomedical Informatics*, 46, 857-868.
- [5] Sue Pelski,. Oracle Multimedia DICOM Developer's Guide, 12c Release 1 (12.1) E17698-09. July 2014.
- [6] Alvez C. y Vecchietti A. “Representación y Recuperación de Imágenes Médicas en Bases de Datos Objeto-Relacionales”. Jornadas Argentinas de Informática e Investigaciones Operativas: 36ª JAIIO– Simposio de Informática y Salud: SIS-2007. Mar del Plata. Agosto de 2007.
- [7] Chuck Murray. Oracle Database. Semantic Technologies Developer's Guide. 11g Release 2 (11.2) E25609-06. January 2014.
- [8] Carlos E. Alvez, Aldo R. Vecchietti. Combining Semantic and Content Based Image Retrieval in ORDBMS. KES'2010, Lecture Notes in Computer Science, 2010, Volume 6277/2010, 44-53. Editors Rossitza Setchi, Ivan Jordanov, Robert J. Howlett, Lakhmi C. Jain. Springer-Verlag Berlin Heidelberg, 2010.
- [9] Borst, W.N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse. PhD. Thesis. University of Twente. 1997.
- [10] Studer, R, V. Benjamins, y D. Fensel. Knowledge Engineering: Principles and Methods. *IEEE Transactions on Data and Knowledge Engineering*, 25, 161-197. 1998.

Búsquedas en Grandes Volúmenes de Datos

Luis Britos, María E. Di Gennaro, Veronica Gil-Costa, Fernando Kasián, Jair Lobos, Verónica Ludueña, Romina Molina, Marcela Printista, Nora Reyes, Patricia Roggero, Guillermo Trabes

LIDIC, Dpto. de Informática, Fac. de Cs. Físico Matemáticas y Naturales, Universidad Nacional de San Luis
{lebritos, mdigena, gvcosta, fkasian, vlud, mprinti, nreyes, proggero}@unsl.edu.ar
jairlobos@gmail.com, mromy00@gmail.com, guillermotrabes@hotmail.com

Edgar Chávez

Centro de Investigación Científica y de Educación Superior de Ensenada, México
elchavez@cicese.mx

Claudia Deco

Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario
deco@fceia.unr.edu.ar

Resumen

En la actualidad los sistemas de información demandan no sólo poder realizar búsquedas eficientes sobre distintos tipos de datos, tales como texto libre, audio, video, secuencias de ADN, etc., sino también poder manejar grandes volúmenes de estos datos. Dada una consulta, el objetivo de un sistema de recuperación de información es obtener lo que podría ser útil o relevante para el usuario, usando una estructura de almacenamiento especialmente diseñada para responderla eficientemente.

Nuestra línea de investigación tiene como principal objetivo desarrollar herramientas para sistemas de información sobre bases de datos masivas, conteniendo datos multimedia, que sean eficientes. Con este fin, se investigan nuevas técnicas que soporten la interacción con el usuario, nuevas estructuras de datos (índices) capaces de manipular eficientemente datos multimedia y que permitan manejar bases de datos masivas de este tipo de datos y se desarrollan nuevas aplicaciones que soporten la recolección y el procesamiento de grandes volúmenes de datos no estructurados.

Palabras Claves: recuperación de información, computación de alto desempeño, grandes bases de datos.

1. Contexto

Esta línea de investigación se encuentra enmarcada dentro del Proyecto Consolidado 3-30114 de la Universidad Nacional de San Luis (UNSL) y en el Programa de Incentivos (código 22/F434): “Tecnologías Avanzadas Aplicadas al Procesamiento de Datos Masivos”, dentro de la línea “Recuperación de Datos e Información”, desarrollada en el Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC) de la UNSL.

En este contexto, se pretende aportar a la incor-

poración de información no estructurada en los procesos de toma de decisiones y resolución de problemas, no considerados en los enfoques clásicos. Por lo tanto, nuestra línea se dedica, principalmente, al diseño de índices eficientes que sirvan de apoyo a sistemas de recuperación de información orientados a conjuntos masivos de datos multimedia. Se espera así contribuir a estos sistemas obteniendo índices más eficientes para memorias jerárquicas, dinámicos, con E/S eficiente, escalables (capaces de manejar grandes volúmenes de datos), considerando técnicas de computación de alto desempeño (HPC).

2. Introducción y Motivación

Con el uso masivo de internet, estamos en presencia de un fenómeno donde la aceleración tanto del crecimiento del volumen de datos capturados y almacenados, como la creciente variación en los tipos de datos requeridos, hace que las técnicas tradicionales para el procesamiento, análisis y obtención de información útil deban ser redefinidas para formular nuevas metodologías de abordaje.

Los sistemas de computación tradicionales hacen uso intensivo de información estructurada; es decir, datos generados con un formato específico. En estos casos, la estructura o formato de esta información puede ser fácilmente interpretada y directamente utilizada por un programa de computadora. Pero el hecho de restringirse al uso de este tipo de información conduce, muchas veces, a representar una visión parcial del problema y dejar fuera información que podría ser relevante para la resolución efectiva del mismo. En este contexto gran parte de la infor-

mación que se requiere para la toma de decisiones y la resolución de problemas de índole más general proviene de información no estructurada.

Habitualmente, se utilizan diferentes métodos de acceso o índices [5] para responder eficientemente a consultas para recuperación de información sobre bases de datos multimedia, principalmente por la gran cantidad de datos con los que se trabaja. Los índices pueden tener distintas características que los hacen indicados para aplicaciones reales: eficientes, dinámicos, escalables, resistentes a la *maldición de la dimensión*, entre otras. Un enfoque prometedor para sistemas de recuperación usando búsqueda por similitud es “la búsqueda basada en contenidos”, la cual usa el dato multimedia mismo. Para calcular la similitud entre dos objetos multimedia, se debe definir una función de distancia. Dicha función mide la disimilitud entre dos objetos.

El concepto de espacios métricos da un marco formal, independiente del dominio de la aplicación, para definir el concepto de búsqueda por similitud. Un espacio métrico está compuesto por un *universo* \mathcal{U} de objetos y una función de distancia $d : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}^+$, que satisface las propiedades que la hacen una métrica. Las consultas por similitud, sobre una *base de datos* $\mathcal{S} \subseteq \mathcal{U}$, son básicamente de dos tipos: *Búsqueda por rango* y *Búsqueda de los k vecinos más cercanos*. La función de similitud (distancia) mide el mínimo esfuerzo (costo) necesario para transformar un objeto en otro. Dependiendo de los tipos de datos multimedia reales el cálculo de dicha función puede ser muy costoso. En particular, para ahorrar cálculos de distancia es importante que dicha distancia satisfaga la desigualdad triangular.

Si la base de datos \mathcal{S} posee n objetos, las consultas se pueden responder llevando a cabo n evaluaciones de distancia. Sin embargo, en la mayoría de las aplicaciones, las distancias son costosas de computar (por ej.: comparación de huellas digitales). En conjuntos masivos de datos la búsqueda secuencial es impráctica y, en general, los repositorios de datos multimedia son grandes volúmenes de datos. Para responder a las consultas con la menor cantidad de cálculos de distancia se debe preprocesar la base de datos para construir un índice. En algunos casos, es probable que la base de datos, el índice, o ambos, no puedan almacenarse en memoria principal. Por lo tanto, para lograr eficiencia, se debe minimizar el número de operaciones de E/S, considerar la jerarquía de memorias y utilizar técnicas paralelas.

Esta propuesta se enfoca en obtener herramientas

de recuperación de información, desarrollando nuevas técnicas y aplicaciones que soporten la interacción con el usuario, diseñando estructuras de datos (índices), capaces de manipular eficientemente grandes volúmenes de datos no estructurados y facilitando la realización de diferentes consultas, de modo de acercar las bases de datos multimedia al nivel de desarrollo de las bases de datos tradicionales.

3. Líneas de Investigación

Se pretende investigar sobre distintos aspectos de los sistemas de recuperación de información multimedia: el diseño de nuevos índices, representaciones que reflejen características de interés de los objetos, distintas consultas sobre estos tipos de bases de datos y eficiencia al considerar grandes volúmenes de datos.

Diseño de Índices

Existen muchos índices para espacios métricos [5]. En su gran mayoría usan la desigualdad triangular para evitar el análisis secuencial de la base de datos. Ésta es la propiedad que permite estimar la distancia entre la consulta q y los objetos de la base de datos, si se han calculado de antemano algunas distancias a objetos distinguidos, y evita calcular las distancias reales desde q a algunos objetos durante una búsqueda. Las distintas técnicas difieren en si esos objetos distinguidos son *pivotes* o *centros*. Si son pivotes se almacenan las distancias de todos los objetos de la base de datos a ellos. Si son centros se particiona el espacio en zonas denominadas *particiones compactas*, por cercanía a los centros y se almacena un radio de cobertura para determinar la zona de cada centro.

Si los objetos de la base de datos se conocen de antemano y luego de construir el índice se realizarán las consultas, los índices se denominan *estáticos*. Por el contrario, si los objetos no se conocen de antemano y el índice se irá creando, preferentemente de manera incremental, a medida que arriben los elementos, permitiendo consultas en cualquier momento, entonces los índices se denominan *dinámicos*. Las estructuras estáticas se benefician al conocer la base de datos completa, ya que pueden seleccionar los mejores puntos de referencia para una estructura de datos determinada. Por el contrario, en las estructuras de datos dinámicas esto no es posible, porque tanto los objetos como las consultas arriban al azar.

Otro aspecto importante para buscar una solución es saber si se puede trabajar en memoria principal

o, por el contrario, si por ser conjuntos de datos masivos se deberá trabajar en otros niveles de la jerarquía de memorias. Además, en este último caso, una manera de lograr eficiencia en las operaciones sobre los índices es aplicando técnicas de computación de alto desempeño y en otros casos mediante la adaptación o diseño de las estructuras que sean concientes de la jerarquía de memorias, minimizando no sólo la cantidad de cálculos de distancia, sino también el número de operaciones de E/S. Otra manera de acelerar la respuesta a una consulta es admitir una respuesta aproximada, permitiendo que la misma sea de menor calidad o menos exacta, pero muy rápida.

Por un lado, nos interesa mejorar el desempeño de índices dinámicos jerárquicos (árboles) para espacios métricos, los cuales se construyen incrementalmente vía inserciones. La raíz del árbol es el primer objeto que llega, y esto se repite recursivamente en cada nivel del árbol. El *Árbol de Aproximación Espacial Distante DiSAT* [4] es un índice muy eficiente en cuanto al número de cálculos de distancias realizados tanto en construcción como en búsquedas. La desventaja del *DiSAT* es que no admite inserciones ni eliminaciones y no es posible construirlo incrementalmente. Sin embargo, una opción para transformarlo en una estructura dinámica es mediante la aplicación de la técnica de Bentley y Saxe [1] que permite lograr dinamismo a partir de una estructura estática, cuando la búsqueda sobre ella cumple con ser un problema que se puede descomponer en partes independientes. Por lo tanto, se está desarrollando un *DiSAT* dinámico usando esta técnica, con un algoritmo eficiente de búsqueda de k -vecinos más cercanos, a pesar de no poder descomponer dicha consulta en búsquedas independientes.

Por otra parte, considerando aplicaciones sobre conjuntos de datos masivos, donde los volúmenes de información con los que se debe trabajar (millones de imágenes en la Web), se hace necesario que los índices sean almacenados en memoria secundaria. En este caso, para hacerlos eficientes, no sólo se debe considerar que en las búsquedas se realice el menor número de cálculos de distancia sino también, dado el costo de las operaciones sobre disco, se efectúe la menor cantidad posible de operaciones de E/S. Hemos diseñado e implementado las siguientes estructuras *DSACL*-tree* y el *DSACL+-tree* [2], optimizadas para memoria secundaria, que demostraron ser competitivas frente a otras de las estructuras conocidas tales como el *M-tree* y *DSA*-tree* y *DSA+-tree* [11]. También, existe una nueva propues-

ta en evaluación (una nueva versión del *DSATCL-tree*) que promete ser aún más eficiente. Además, es posible mejorar su desempeño mediante la aplicación de técnicas paralelas. Por lo tanto, se buscará aplicar y comparar distintas estrategias de paralelización con el fin de determinar la más adecuada.

Por otro lado, tomando como base al índice para búsquedas aproximadas *Lista de Permutaciones Agrupadas* (LPA), que combina un algoritmo basado en *Permutaciones* con una *Lista de Clusters* [6], se ha propuesto una nueva versión de la LPA que permite realizar búsquedas por similitud aproximadas sobre conjuntos de datos masivos [13]. Esta nueva versión de la LPA es conciente que trabaja en memoria secundaria y no sólo considera minimizar los costos en cantidad de distancias calculadas, sino también en cantidad y tipo de operaciones de E/S.

Consultas sobre Bases de Datos Multimedia

Las operaciones tradicionales sobre bases de datos multimedia son las búsquedas por rango o de k -vecinos más cercanos. Sin embargo, existen otras operaciones de interés como las distintas variantes del *join* por similitud. Para estas operaciones se consideran dos bases de datos A y B , ambas subconjuntos del mismo universo del espacio métrico \mathcal{U} . El resultado de cualquier operación de *join* por similitud entre A y B obtiene el conjunto de pares formados por un objeto de A y otro de B , tales que entre ellos se satisface el predicado de similitud Φ considerado. Las variantes más conocidas son: el *join* por rango, el *join* de k -vecinos más cercanos y el *join* de k pares de vecinos más cercanos; entre otras.

Formalmente, dadas $A, B \subseteq \mathcal{U}$, se define el *join por similitud* entre A y B ($A \bowtie_{\Phi} B$) como el conjunto de todos los pares (x, y) , donde $x \in A$ e $y \in B$; es decir, $(x, y) \in A \times B$, tal que $\Phi(x, y)$ es verdadero (se satisface el criterio de similitud Φ entre x e y). Al resolver el *join* por similitud es posible que ambas, una o ninguna de la bases de datos posean un índice; o que ambas bases de datos se indexen conjuntamente con un índice diseñado para el *join*. Calcular cualquiera de las variantes del *join* por similitud de manera exacta es muy costoso [12], así vale la pena analizar posibilidades de obtener una respuesta aproximada al *join*, más rápidamente, aunque siempre buscando buena calidad en la respuesta.

PostgreSQL es el primer sistema de base de datos que permite realizar consultas por similitud sobre algunos atributos, particularmente indexa para búsquedas de k -vecinos más cercanos (índices *KNN-GiST*). Estos índices pueden ser usados sobre tex-

to, comparación de ubicación geoespacial, etc. Sin embargo, los índices *K-NN GiST* proveen plantillas sólo para índices con estructura de *árbol balanceado* (*B-tree*, *R-tree*), pero el “balance” no siempre es bueno para los índices que se utilizan en búsquedas por similitud [3]. Además, no se dispone de este tipo de consultas para todo tipo de datos métricos. Así, es importante proveer un DBMS para bases de datos métricas que maneje todos los posibles datos métricos y las operaciones de interés sobre ellos [7].

Simulaciones Paralelas Aproximadas para Sistemas de Gran Escala

Los *motores de búsqueda Web* (WSE) son sistemas complejos y sumamente optimizados que operan sobre clusters de procesadores. Los WSE gestionan cargas de trabajo altamente dinámicas. La evaluación del desempeño de estos WSE es de suma importancia para la implementación y funcionamiento de los centros de datos y para labores de investigación.

La simulación secuencial de eventos discretos y la simulación orientada a procesos han demostrado ser útiles en estudios de evaluación de desempeño en modelos de pequeña escala. Sin embargo, para el estudio de modelos de gran escala se utiliza simulación paralela por eventos discretos, ya que permite satisfacer requerimientos críticos, como tiempos de espera de resultados de la simulación o de consumo de memoria. Existen dos enfoques principales para efectuar simulaciones paralelas: el enfoque optimista y el conservativo. Ambos tienen como objetivo garantizar el cumplimiento de la causalidad de eventos. Los protocolos conservativos resultan inviables en simulaciones de grandes WSE, que se caracterizan porque los objetos simulados presentan patrones de comunicación aleatorios en que todos los elementos se comunican con todos, o que pueden incluir implementaciones reales como cachés de resultados, y en que los incrementos de tiempo de los eventos pueden presentar grandes diferencias entre ellos.

Dentro del proyecto de investigación descrito en este trabajo se propone desarrollar y evaluar un esquema optimista de simulación paralela semi-síncrona y aproximado [10, 14], que reduzca los tiempos de ejecución y el consumo de memoria de simulaciones optimistas. El objetivo es que los resultados estimados por simulaciones aproximadas, que violan la restricción de causalidad, presenten alta precisión respecto a resultados de simulaciones secuenciales. El nivel de optimismo se podría gestionar eficientemente a lo largo de la simulación, y

se ajustaría automáticamente a características propias de la simulación. El esquema a abordar utiliza un algoritmo para balancear la carga de los procesadores que ejecutan la simulación paralela.

Por lo tanto, se busca: (a) proponer y evaluar un nuevo enfoque de simulación paralela aproximado para sistemas de búsqueda Web de gran escala y (b) utilizar el framework PCD++ [8] basado en DEVS para diseñar y desarrollar el nuevo enfoque.

Sistema CBIR con Plataforma de Base FPGA

Con el crecimiento constante de los datos que se desarrollan en diferentes aplicaciones, pero con mayor dimensión en la Web, estas nuevas aplicaciones exigen el uso de datos más complejos que sólo texto sin formato. En estos casos, los espacios métricos han demostrado ser útiles y prácticos para realizar una búsqueda de similitud en grandes colecciones de objetos complejos. En este caso, las consultas son objetos del mismo tipo de los almacenados en la base de datos cuando, por ejemplo, uno está interesado en la recuperación de los k objetos más similares a una consulta dada. La similitud entre dos objetos se calcula mediante una función de distancia dependiente de la aplicación, que generalmente suele ser costosa para calcular en términos computacionales.

Las aplicaciones de recuperación de imágenes basada en contenido visual también se conocen como sistemas de recuperación de imágenes basada contenido (CBIR). Los sistemas CBIR se pueden modelar con espacios métricos e implican básicamente dos operaciones: (a) cálculos de descriptores visuales que caracterizan y describen el contenido de una imagen (los vectores característicos o descriptor); y (b) la medida de similitud para seleccionar imágenes candidatas de la base de datos similares a una consulta (para el proceso de búsqueda).

Dentro de este proyecto se propone implementar un sistema CBIR sobre una plataforma SoC basado en FPGA para acelerar la búsqueda de similitud mediante el aprovechamiento de los beneficios de las herramientas de síntesis de alto nivel. El objetivo es reducir el tiempo de ejecución requerido para construir el vector de descriptor de una imagen, cuyo proceso representa uno de los cuellos de botella en los sistemas de CBIR [15]. Así, se tiene por objetivo: (1) proponer, diseñar y evaluar un sistema CBIR sobre una plataforma SoC basada en FPGAs y (2) tomar ventaja de las propiedades de área, consumo de energía y velocidad de procesamiento provista por las nuevas plataformas SoCs.

4. Resultados

Se implementaron el *DSACL*-tree* y *DSACL+-tree*, que trabajan con grandes volúmenes de datos, diseñadas para memoria secundaria y que mostraron ser competitivas contra otras estructuras diseñadas para tal fin [2]. Se espera lograr una implementación paralela eficiente de estos índices. Se ha obtenido una versión para memoria secundaria de la *LPA* [13]. Se ha completado el análisis del *DiSAT* [4] y se espera proponer una versión dinámica del mismo, que mantenga su eficiencia. Se está trabajando en la extensión de *PostgreSQL* que incluya más consultas por similitud y sobre distintos tipos de datos.

Habiendo previamente desarrollado un conjunto de aplicaciones de comparación para determinar las operaciones más relevantes y costosas de la plataforma *S4* [9], se ha propuesto y evaluado un simulador paralelo asíncrono, diseñado para procesamiento de “streams” distribuidos sobre dicha plataforma [14].

5. Formación de Recursos

En esta línea se están realizando los siguientes trabajos de formación en Ciencias de la Computación:

Tesis de Doctorado: (1) “Planeación de Capacidad en Centros de Datos para Sistemas Escalables para la Web” (con beca de posgrado de CONICET), (2) “Planeación de Capacidad para Motores de Búsqueda Web” y (3) “Optimización Dinámica de Funcionamiento de Motor de Búsqueda con Máquinas de Aprendizaje y Transformada Discreta de Fourier”.

Tesis de Maestría: (1) “Estructuras Eficientes sobre Datos Masivos para Búsquedas en Espacios Métricos”, (2) “Simulación Paralela Aproximada sobre *S4* para Motores de Búsqueda en la Web”, (3) “Recuperación de Imágenes sobre Plataformas de Sistemas de Cómputo de Alta Productividad” y (4) un “Sistema Administrador para Bases de Datos Métricas”.

Trabajo Final de Licenciatura: “Aplicación de Multi-BSP para la Estimación de Costo de Algoritmos sobre Plataformas Multi-core”.

Referencias

- [1] J. Bentley and J. Saxe. Decomposable searching problems: Static-to-dynamic transformation. *J. Algorithms*, 1(4):301–358, 1980.
- [2] L. Britos, A. M. Printista, and N. Reyes. Dynamic spatial approximation trees with clusters for secondary memory. In *XVI CACIC Selected Papers*, Computer Science & Technology Series, 205–215. Editorial UNLP, 2011.
- [3] E. Chávez, V. Ludueña, and N. Reyes. Revisiting the VP-forest: Unbalance to improve the performance. In *Procs. of JCC*, page 26, 2008.
- [4] E. Chávez, V. Ludueña, N. Reyes, and P. Roggero. Faster proximity searching with the distal {SAT}. *Information Systems*, pages–, 2016. In Press, Available online.
- [5] E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín. Searching in metric spaces. *ACM*, 33(3):273–321, Sept. 2001.
- [6] K. Figueroa and R. Paredes. List of clustered permutations for proximity searching. In *Procs. of the 6th SISAP.*, vol. 8199, LNCS, 50–58. Springer, 2013.
- [7] F. Kasián and N. Reyes. Búsquedas por similitud en PostgreSQL. In *Actas del XVIII CACiC*, 1098–1107, 2012.
- [8] Q. Liu and G. Wainer. Lightweight time warp - a novel protocol for parallel optimistic simulation of large-scale devs and cell-devs models. In *12th IEEE/ACM Int. Symp. on DS-RT.*, 131–138, Oct. 2008.
- [9] J. Lobos, V. Gil-Costa, and M. Marin. Benchmark applications for stream processing profiling. In *Procs. of JCC*, Talca, Chile, 2014.
- [10] M. Marin, V. Gil-Costa, C. Bonacic, and R. Solar. Approximate parallel simulation of web search engines. In *Procs. of 1st ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*, 189–200, 2013. ACM.
- [11] G. Navarro and N. Reyes. Dynamic spatial approximation trees for massive data. *Procs. of 2nd SISAP*, 81–88. IEEE Comp. Soc., 2009.
- [12] R. Paredes and N. Reyes. Solving similarity joins and range queries in metric spaces with the list of twin clusters. *JDA*, 7:18–35, 2009.
- [13] P. Roggero, N. Reyes, K. Figueroa, and R. Paredes. List of clustered permutations in secondary memory for proximity searching. *JCS&T*, 15(2):107 – 113, Nov. 2015.
- [14] E. Tapia, V. Gil-Costa, and M. Marin. Evaluation of a parallel simulation algorithm for the *s4* stream processing platform. In *Procs. of JCC*, Santiago, Chile, 2015.
- [15] Y. Zhang. *Advances in Face Image Analysis: Techniques and Technologies*. IGI Global, USA, 1st edition, 2010.

Minería de Datos Aplicada a Datos Masivos

**Anabella De Battista, Patricia Cristaldo, Lautaro Ramos,
Juan Pablo Nuñez, Soledad Retamar, Daniel Bouzenard**

Departamento Ingeniería en Sistemas de Información

Fac. Reg. Concepción del Uruguay

Universidad Tecnológica Nacional

Entre Ríos, Argentina

{debattistaa, cristaldop, ramosl, nunezjp, retamars, bouzenardd}@frcu.utn.edu.ar

Norma Edith Herrera

Departamento de Informática

Univ. Nac. de San Luis

San Luis, Argentina

nherrera@unsl.edu.ar

Resumen

Las grandes cantidades de datos que se producen en la actualidad, sumadas a su heterogeneidad, hacen que las herramientas tradicionales de análisis de datos no resulten adecuadas para su recopilación, almacenamiento, gestión y análisis. En este contexto se comienza a hablar del término *Big Data*, haciendo referencia a características como gran volumen, velocidad y variedad de producción de los datos, y a las herramientas que se utilizan para encontrar valor en las mismas. La posibilidad de hallar patrones y tendencias en estas grandes cantidades de datos impacta directamente en la toma de decisiones en áreas tan diversas como salud, genética, agro, predicciones climáticas, redes sociales, marketing, finanzas, educación, entre otras. Otro aspecto de interés en este tipo de análisis, es la aplicación de metodologías de gestión de proyectos de enfoque ágil en los proyectos de

minería de datos, en este caso, se aplicarán metodologías específicas con el objetivo de comparar características y restricciones de cada una. En este artículo se presentan los tópicos de interés del proyecto *Minería de Datos: su aplicación a repositorios de datos masivos*.

Palabras Claves: *Big Data*, minería de datos, clustering, agrupamiento, gestión de proyectos, CRISPDM.

1. Contexto

El presente trabajo se desarrolla en el ámbito del proyecto *Minería de Datos: su aplicación a repositorios de datos masivos* (EITU0003781TC) del Grupo de Investigación en Bases de Datos, perteneciente al Departamento Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, F. R. Concepción del Uruguay.

2. Introducción

Actualmente se producen diariamente grandes volúmenes de datos de diversos tipos (e.g., textos, imágenes, audio, videos) y desde los más variados orígenes (e.g., web, GPS, redes sociales, sensores). Se prevé que en los próximos años las aplicaciones *Internet de las cosas* aumentarán el volumen de datos a un nivel sin precedentes. En este contexto, surge el término *Big Data* referido a conjuntos de datos cuyo tamaño supera la capacidad de las herramientas tradicionales de bases de datos de recopilar, almacenar, gestionar y analizar la información. En general se habla de *Big Data* o Análisis de *Big Data* como sinónimos, ya que no sólo se desea hacer referencia a la gran cantidad y complejidad de los datos, sino también a las herramientas utilizadas para procesarlos y extraer conocimiento útil de los mismos.

Algunas definiciones indican que *Big Data* puede definirse a partir de las siguientes características [1]:

- Volumen: órdenes superiores a Terabytes de datos
- Variedad: distintos tipos de datos provenientes de diversas fuentes que pueden organizarse tanto en forma estructurada como no estructurada.
- Velocidad: referido a la velocidad de generación de los datos o a la rapidez con la que se generan y procesan los datos.
- Variabilidad: referido a la inconsistencia que pueden presentar los datos en ocasiones, dificultando las tareas de análisis.
- Valor: gracias a la posibilidad de tomar decisiones al responder preguntas que antes no era posible, ofrece a la organización una ventaja estratégica.

Las actividades de la comunidad científica y profesional han cambiado debido al surgimiento de estos grandes repositorios de datos, ya

que se requieren nuevas estrategias para su almacenamiento, tratamiento, distribución y análisis, porque no sólo se cuenta con una gran cantidad de datos sino que además su complejidad es creciente. La posibilidad de hallar patrones y tendencias en estas grandes cantidades de datos impacta directamente en la toma de decisiones en áreas tan diversas como salud, genética, agro, predicciones climáticas, redes sociales, marketing, finanzas, educación, entre otras. Es por esta razón que se comienza a trabajar en nuevas herramientas, tecnologías, métodos y sistemas requeridos para manejar grandes conjuntos de datos distribuidos, heterogéneos, no estructurados, diversos y complejos. La detección de agrupamientos en repositorios de datos grandes y complejos es una de las actividades más relevantes en análisis de información. Por ejemplo, dada una base de datos de imágenes satelitales de varias decenas de terabytes: ¿Es posible encontrar regiones con el objetivo de identificar selvas naturales, deforestación o reforestación, o identificar terreno cultivado y el tipo de cultivo? ¿Puede hacerse automáticamente? La respuesta a ambos interrogantes es "sí". En la actualidad ese análisis puede realizarse en pocos minutos con muy alta precisión [2]. Por otro lado, la clasificación de datos complejos no puede realizarse con las herramientas tradicionales de análisis de datos, por lo que han surgido nuevos algoritmos especialmente diseñados para el análisis de datos masivos. La Minería de Datos involucra e integra técnicas de diferentes disciplinas tales como tecnologías de bases de datos y data warehouse, estadísticas, aprendizaje de máquinas, computación de alta performance, reconocimiento de patrones, redes neuronales, visualización de datos, recuperación de información, procesamiento de imágenes y señales, y análisis de datos espaciales o temporales. En este proyecto se estudiarán procesos de Minería de Datos desde una perspectiva de bases de datos, con enfoque en técnicas eficientes y escalables.

3. Líneas de Investigación, Desarrollo e Innovación

La línea de trabajo principal de nuestro proyecto de investigación es el estudio de técnicas de Minería de Datos aplicables a repositorios de datos masivos, atendiendo principalmente a su eficiencia y escalabilidad. En particular se realizará el estudio, análisis y comparación del funcionamiento de algoritmos de clustering y de clasificación aplicables a datos masivos, realizando pruebas en distintos modelos de bases de datos (espaciales, temporales, espacios métricos) para posteriormente proponer mejoras a los algoritmos existentes o bien, nuevos algoritmos [3]. Además se pretende desarrollar aplicaciones que implementen algoritmos de clustering o de clasificación, para el tratamiento de repositorios de datos de alta complejidad.

En la gestión de los proyectos específicos de Minería de Datos que se emprendan en este proyecto de investigación se emplearán distintas metodologías de gestión de proyectos de enfoque ágil y se realizará una comparación de características y restricciones de cada metodología, en relación a los factores críticos de éxito y las razones de fracaso en la gestión de proyectos de enfoque ágil [4, 5].

3.1. Técnicas de Minería de Datos Aplicables a Datos Masivos

La gran cantidad de datos que actualmente generan y almacenan aplicaciones de diversas áreas, está en continuo crecimiento, no sólo desde el punto de vista de objetos y atributos, sino en la complejidad de los atributos que describen a cada objeto [2]. La masividad de estos datos ha superado nuestra capacidad de procesar, almacenar adecuadamente, analizar y entender estos grandes repositorios. Es de gran interés para las organizaciones propietarias de datos masivos poder extraer conocimiento de ellos, convirtiéndolos en recursos útiles para

la toma de decisiones en lugar de sólo mantenerlos resguardados en discos de computadoras, sin ser accedidos nunca. En este contexto surge el área de investigación conocida como Descubrimiento de Conocimiento en Bases de Datos (KDD por sus siglas en inglés), que se define como el proceso no trivial de identificar patrones válidos, desconocidos, potencialmente útiles y comprensibles en los datos. El proceso de KDD consta de una secuencia iterativa de etapas: integración y recopilación de datos; selección, limpieza y transformación de datos; minería de datos; evaluación; difusión, uso y monitorización de modelos. Actividades comunes dentro del proceso de KDD son clustering, clasificación y etiquetado, identificación de errores de medición, detección de outliers, inferencia de reglas de asociación y datos ausentes, y reducción de la dimensionalidad. KDD es un proceso complejo que tiene un alto costo computacional, producto de explorar varios elementos en distintas combinaciones para lograr el conocimiento deseado [2]. En las bases de datos tradicionales los datos se representan mediante atributos numéricos o categorizados en una tabla, donde cada tupla representa un elemento del conjunto. El desempeño de los algoritmos de análisis de datos en general depende del número de elementos en el conjunto, o de la cantidad de atributos en la tabla, y de las distintas formas en que interactúan tuplas y atributos. En un contexto de *Big Data* las técnicas tradicionales de KDD no son suficientes por la complejidad y magnitud de los repositorios de datos. La fase de Minería de Datos es la más característica del KDD por eso con frecuencia se utiliza esta fase para nombrar todo el proceso. Su objetivo es producir nuevo conocimiento que pueda ser útil al usuario. Esto se realiza construyendo un modelo basado en los datos recopilados a tal fin. El modelo describe los patrones y relaciones entre los datos que pueden utilizarse para: realizar predicciones, entender mejor los datos, explicar comportamientos pasados. Para

ello se deben tomar ciertas decisiones antes de comenzar el proceso: determinar qué tipo de tarea de minería es el más apropiado (por ejemplo, se podría utilizar clasificación para identificar en una universidad que alumnos abandonarían sus estudios); seleccionar el tipo de modelo (para una tarea de clasificación se podría utilizar un árbol de decisión para obtener un modelo en forma de reglas); seleccionar el algoritmo de minería que resuelva la tarea y obtenga el tipo de modelo buscado [6].

Existen distintos tipos de tareas dentro de la minería de datos, cada una con sus propios requisitos y que retornan información de distintos tipos. Las tareas se clasifican en:

- Predictivas: a su vez de distinguen las tareas de clasificación (técnicas: árboles de decisión, reglas de asociación, redes bayesianas, redes neuronales, SVM) y regresión.
- Descriptivas: encontramos las siguientes tareas: reglas de asociación, correlaciones, clustering, detección de anomalías.

La Minería de Datos puede aplicarse sobre cualquier tipo de repositorio de datos (BD Relacionales, Data Warehouses, BD Transaccionales, en la Web y en Sistemas de BD avanzados como BD Objeto-relacionales, BD Espaciales, BD Temporales, BD de Textos, BD Multimedia). Las técnicas a aplicar varían de acuerdo a cada tipo de repositorio [7].

3.2. Metodologías de Gestión de Proyectos de Minería de Datos Masivos

En los últimos años se han propuesto en el área de KDD nuevas técnicas que puedan gestionar Datos Masivos, pero también se han estudiado las metodologías que permiten acceder al nuevo conocimiento que se pretende encontrar. Las metodologías nos permiten llevar a cabo el proceso de minería de datos

en forma sistemática y no trivial, definiendo además de las fases del proyecto, las tareas a realizar y cómo llevarlas a cabo. Según la filosofía de desarrollo, tanto las guías como las metodologías, se pueden clasificar según dos enfoques: tradicionales, que se basan en una fuerte planificación durante toda la gestión del proyecto y un ciclo de vida más lineal; y metodologías ágiles, en las que la gestión del proyecto es incremental, cooperativo, ampliamente adaptable y abiertas al cambio. El término KDD fue acuñado en el año 1996 y constituyó el primer modelo aceptado en la comunidad científica que establece las etapas principales de un proyecto de explotación de información. En su versión completa, KDD está formado por nueve etapas [8, 9]. A partir del año 2000, con el gran crecimiento en el área de minería de datos, surgen tres nuevos modelos que plantean un enfoque sistemático para llevar a cabo el proceso [10]: SEMMA, CRISP-DM y Catalyst (conocida como P3TQ). Algunas de las metodologías profundizan en mayor detalle sobre las tareas y actividades a ejecutar en cada etapa del proceso de minería de datos (como CRISP-DM), mientras que otras proveen sólo una guía general del trabajo a realizar en cada fase (como SEMMA). En este proyecto se estudiará la aplicación de metodologías ágiles en proyectos de minería de datos masivos.

4. Resultados y Objetivos

Con este proyecto se espera proponer modificaciones o mejoras a los algoritmos de clustering o de clasificación existentes para datos masivos, o proponer nuevos algoritmos, además del desarrollo de aplicaciones que implementen este tipo de algoritmos en el tratamiento de repositorios de datos masivos. A partir de la aplicación y comparación de metodologías de gestión de proyectos de enfoque ágil a estos proyectos de descubrimiento de conocimiento, se espera poder determinar

cual es la metodología mas adecuada a aplicar en proyectos de estas características.

5. Formación de Recursos Humanos

Este proyecto da inicio a una nueva línea de investigación dentro del Grupo de investigación en Bases de Datos de la Fac. Reg. Concepción del Uruguay de la U.T.N.. La Directora y la codirectora del proyecto, también lo son de las carreras de posgrado Especialización y Maestría en Ciencias de la Computación con orientación en Bases de Datos que se dictan en esta Facultad. Tres de los investigadores del proyecto están desarrollando sus tesis de maestría luego de haber finalizado la cursada de dicho posgrado. Se cuenta con un becario graduado, que está iniciando su camino en la investigación y prevé la realización de un posgrado en el área temática del proyecto. Una de las integrantes del grupo está desarrollando su Tesis Doctoral sobre la temática de indexación en memoria secundaria de bases de datos textuales, tema íntimamente relacionado a las líneas de estudio de este grupo. El grupo cuenta en la actualidad con dos becarios alumnos de la carrera Ingeniería en Sistemas de Información que inician su formación en la investigación.

Referencias

- [1] Fan Wei and Bifet Albert. Mining big data: Current status, and forecast to the future. *SIGKDD Explor. Newsl.*, 14(2):1–5, apr 2013.
- [2] Robson Leonardo Ferreira Cordeiro, Christos Faloutsos, and Caetano Traina Junior. *Data Mining in Large Sets of Complex Data*. SpringerBriefs in Computer Science. Springer-Verlag London, 2013.
- [3] Larose Daniel T. *Discovering Knowledge in Data: An Introduction to Data Mining*. Wiley-Interscience, 2004.
- [4] Tsun Chow and Dac-Buu Cao. A survey study of critical success factors in agile software projects. *Journal of Systems and Software*, 81(6):961 – 971, 2008. Agile Product Line Engineering.
- [5] Gabriella Cserhati and Lajos Szabo. The relationship between success criteria and success factors in organisational event projects. *International Journal of Project Management*, 32(4):613 – 624, 2014.
- [6] J.H. Orallo, M.J.R. Quintana, and C.F. Ramirez. *Introducción a la minería de datos*. Fuera de colección Out of series. Editorial Alhambra S. A. (SP), 2004.
- [7] J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2011.
- [8] H. Dai, R. Srikant, and C. Zhang. *Advances in Knowledge Discovery and Data Mining: 8th Pacific-Asia Conference, PAKDD 2004, Sydney, Australia, May 26-28, 2004, Proceedings*. Number v. 8 in Lecture Notes in Artificial Intelligence. Springer, 2004.
- [9] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. The kdd process for extracting useful knowledge from volumes of data. *Commun. ACM*, 39(11):27–34, nov 1996.
- [10] Gonzalo Mariscal, Oscar Marban, and Covadonga Fernandez. A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, 25:137–166, 6 2010.

NoSql en sistemas distribuidos sobre Cluster Hadoop

Adriana E.Martin¹, Susana B. Chavez², Nelson R. Rodríguez³, María A. Murazzo⁴

Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.
Complejo Universitario Islas Malvinas, Av. I. de la Roza 590 (O), CP: 5402 Rivadavia,
San Juan. Tel:0264 4234129

¹arianamartinsj@gmail.com ²schavez@iinfo.unsj.edu.ar ³nelson@iinfo.unsj.edu.ar
⁴marite@unsj-cuim.edu.ar

Resumen

Con el gran avance de los sistemas distribuidos en la web y las tecnologías informáticas distribuidas como Cluster y Cloud, el software como servicio (SaaS), los servicios en el Cloud y los constantes requerimientos de procesamiento y análisis a gran escala de enormes cantidades de datos, los sistemas tradicionales de base de datos son insuficientes. Las Bases de Datos NoSQL llenan una importante carencia de las bases de datos relacionales en cuanto a la capacidad que estas tienen en escalabilidad, distribución y manejo de datos no estructurados. Estas 3 características son cada día más relevantes debido precisamente al avance de Cloud Computing, y a los múltiples y diversos servicios cuyo crecimiento y replicación distribuida son extremadamente necesarios.

El framework Hadoop Map/Reduce permiten realizar procesamiento distribuido de grandes volúmenes de datos. La naturaleza distribuida de Cassandra coincide muy bien con la naturaleza distribuida de MapReduce, para ejecutar consultas sobre datos que abarcan múltiples nodos. La sencillez y flexibilidad de Cassandra, su lenguaje

de consulta (CQL) y el soporte del controlador múltiple como la capa de almacenamiento de datos para MapReduce, tanto para la entrada y salida de datos; permiten sobreponerse a las limitaciones del Sistema de Archivo de Hadoop (HDFS)

Palabras Claves: Computación distribuida, Cloud Computing, Nosql, MapReduce

Contexto

El presente trabajo se encuadra dentro del área *Base de Datos y Minería de Datos*, y se enmarca dentro del proyecto de investigación: **Evaluación de arquitecturas distribuidas de commodity basadas en software libre**, el cual se ha presentado en una nueva convocatoria, y cuyo resultados se espera que estén para abril del presente año. Tendrá como unidades ejecutoras al Departamento e Instituto de Informática de la FCEFYN de la UNSJ.

Introducción

La computación distribuida es un modelo destinado a resolver problemas

de cómputo masivo utilizando un gran número de computadoras organizadas sobre una infraestructura de telecomunicaciones distribuida. El objetivo principal de la computación distribuida es contar con un entorno de entidades distribuidas que cooperen para resolver un problema que no puede ser resuelto individualmente. La variedad de estos sistemas pueden incluir servidores de aplicaciones, cloud privados, pequeños centros de datos, cluster de cómputo y cluster para almacenamiento y búsqueda de datos. También es posible la construcción de arquitecturas híbridas que permitan montar cluster como servicio (CaaS), facilitando hacer el mejor aprovechamiento de los recursos virtualizados en el cloud, con el objeto de utilizar ecosistemas, tales como: Hadoop MapReduce para realizar un procesamiento distribuido de grandes volúmenes de datos.

Con el fin de gestionar estas grandes cantidades de datos, que no tienen en general una estructura determinada, surgen las Bases de Datos NoSQL[1].

Apache Cassandra es una base de datos NoSQL distribuida y de código abierto, cuya principal característica es que fusiona Dynamo, de Amazon con BigTabla, de Google, siendo ambas implementaciones de código cerrado. Está basada en un modelo de almacenamiento de «clave-valor», escrita en Java. Permite almacenar grandes volúmenes de datos en forma distribuida. Su objetivo principal es la escalabilidad lineal y la disponibilidad. [2].

La arquitectura distribuida de Cassandra está basada en una serie de nodos iguales que se comunican con un protocolo P2P (Peer-to-peer). En este protocolo, la Red está descentralizada, no tiene clientes ni servidores fijos, sino

que tiene una serie de nodos que se comportan simultáneamente como clientes y servidores de los demás nodos de la red. Cada nodo puede iniciar, detener o completar una transacción compatible. Contrasta con el modelo cliente-servidor. Esto lleva a una comunicación en la que la redundancia es máxima.

En las versiones iniciales utilizaba un API propia para poder acceder a la base de datos. En los últimos tiempos están apostando por un lenguaje denominado CQL (Cassandra Query Language) que posee una sintaxis similar a SQL aunque con muchas menos funcionalidades. Esto hace que iniciarse en el uso de la misma sea más sencillo. Permite acceder en Java desde JDBC.

“Apache Cassandra se ha convertido en una de las bases de datos NoSQL más utilizados del mundo y sirve como columna vertebral de algunas aplicaciones muy populares hoy en día”, Por ejemplo, lo usa Twitter para su plataforma. Su objetivo principal es la escalabilidad lineal y la disponibilidad. Permite el uso de Hadoop para implementar Map-Reduce, ya que Hadoop puede trabajar directamente con cualquier sistema de archivos distribuido [3]

Está disponible en Linux rpm, Deb, y también existe, un instalador MSI de Windows, y Mac OS X binary.

MapReduce es fundamental en las bases de datos NoSQL para permitir la utilización de funciones de agregación de datos, ya que al carecer de esquema son mucho más complicadas que en las bases de datos relacionales clásicas RDBMS (Relational Database Management System).

MapReduce es un paradigma de programación para el procesamiento y manejo de grandes conjuntos de datos. Se divide las solicitudes en solicitudes más pequeñas que se envían a muchos servidores pequeños para ser procesados de una manera paralela. Como resultado, puede procesar grandes conjuntos de datos muy rápidamente. Común / Core es un paquete que contiene las bibliotecas y los servicios públicos de apoyo a los módulos de Hadoop. El hilo es una plataforma de gestión de recursos incluidos para la gestión de los recursos de computación y programación de tareas de Hadoop.

Hadoop es un framework de Apache de código abierto basado en Java, se ejecuta sobre la máquina virtual de Java (necesario JRE 1.6 o superior) para procesar y consultar grandes volúmenes de datos en grandes clusters de hardware commodity.

La arquitectura Hadoop esta formada por los siguientes módulos:

1. Hadoop Common: Utilidades comunes en las que se apoyan los otros módulos. Proporciona abstracciones a nivel de sistema de archivos o sistema operativo. Contiene los archivos .jar (Java ARchive) y scripts necesarios para iniciar Hadoop.

– Hadoop Distributed File System (HDFS): Sistema de archivos distribuido, que almacena datos en las máquinas commodity, que proporciona acceso de alto rendimiento a datos.

– Hadoop YARN: Plataforma de administración responsable de la gestión de los recursos en el cluster y de los medios para el desarrollo de aplicaciones de los usuarios.

– Hadoop MapReduce: Modelo de programación para el procesamiento de grandes volúmenes de datos.[4]

Hadoop utiliza una arquitectura escalable que hace uso de servidores básicos configurados como un conjunto, donde cada servidor posee unidades de disco internas de bajo costo. Los datos de Hadoop se dividen en bloques y se extienden a lo largo de un clúster. Una vez que esto sucede, las tareas de MapReduce pueden llevarse a cabo en los subconjuntos más pequeños de datos que pueden hacer que un muy gran conjunto de datos en general, cumpliendo así el tipo de escalabilidad necesaria para el procesamiento de grandes volúmenes de datos.[5]

Los cluster Hadoop son conocidos por aumentar la velocidad de las aplicaciones de análisis de datos. También son altamente escalable: Si la potencia de procesamiento está desbordado por volúmenes de datos cada vez mayor, nodos adicionales del clúster se pueden agregar para aumentar el rendimiento. Los cluster Hadoop también son altamente resistentes a fallos, ya que cada pieza de datos se copia en otros nodos del clúster, lo que asegura que los datos no se pierden si falla un nodo.

Un clúster Hadoop esta formado de la siguiente manera:

- Un nodo maestro. En el nodo maestro encontramos un JobTracker (rastreador de trabajos), un TaskTracker (rastreador de tareas), un NameNode (nodo de nombres) y un DataNode (nodo de datos).

– varios nodos esclavos. Un nodo esclavo actúa tanto de DataNode como de TaskTracker normalmente.

Hadoop Distributed File System (HDFS) es un sistema de archivos distribuido que se parece a cualquier otro sistema de archivos, excepto que cuando se mueve un archivo en HDFS, este archivo se divide en muchos archivos pequeños, cada uno de esos archivos se replica y se almacena en (por lo general, puede personalizar) 3 servidores para las limitaciones de tolerancia a fallos.

En general, esta estrategia de divide y vencerás de procesamiento de datos no es nada nuevo, pero la combinación de HDFS, siendo el software de código abierto (que supera la necesidad de soluciones de almacenamiento especializados de alto precio), y su capacidad para llevar a cabo un cierto grado de redundancia y tolerancia a fallos automático, hacen que sea popular para las empresas modernas que buscan soluciones de análisis de proceso por lotes

Líneas de Investigación y desarrollo

Las diferencias entre los sistemas distribuidos es que plantean numerosos desafíos a resolver como son la configuración adecuada, optimización de los recursos, balance de carga, uso eficiente de la red, distribución adecuada de los datos, configuración de frameworks, elección del tipo de hipervisor, uso y configuración del mismo, opciones de virtualización, etc. El objetivo del presente proyecto es evaluar las distintas variantes de sistemas distribuidos de recursos limitados, y proponer en función de los resultados las opciones más convenientes para cada caso. Vale aclarar, que cuando se decide contar con una solución basada en cloud es de suma importancia analizar cual despliegue se usará, pues esta elección impacta directamente en la performance

de la solución. Hay que recordar que cualquier elección que implique el uso intensivo de comunicaciones degrada el desempeño por los retardos incurridos en el proceso de comunicación en si [6] [7].

También es importante destacar que es posible la construcción de arquitecturas híbridas que permitan montar cluster como servicio (CaaS), para lograr el mejor aprovechamiento de los recursos en el cloud, con el objeto de utilizar Hadoop MapReduce para realizar un procesamiento distribuido de grandes volúmenes de datos

Resultados Obtenidos

A pesar de que esta línea de investigación recién ha sido presentada para los años 2016 y 2017. Durante los últimos cinco años se trabajó en proyectos sobre Cloud Computing y en particular durante los últimos dos años sobre Cloud híbridos. La experiencia sobre los Cloud privados, junto con líneas de investigación anteriores [7], impulsó esta línea de investigación. El grupo ha realizado nueve publicaciones en el área durante el último año: tres trabajos en el WICC 2015, un trabajo en el CACIC 2015, dos trabajos en las Jornadas de Cloud Computing, además se realizaron tres publicaciones en revistas científicas.

Se han aprobado tres tesinas de grado y un trabajo de especialización.

Objetivo

En este proyecto se enfocaran las investigaciones en los sistemas de cómputo distribuidos, los cuales permiten realizar de manera más eficiente tareas de computación de alta prestaciones basadas en el paradigma de memoria distribuida. Ejemplos de Arquitecturas que soportan este tipo de

sistemas distribuidos son los cluster y el cloud computing.

En particular este trabajo tiene como objetivo instalar una base de datos Nosql, en particular Cassandra, sobre un cluster montado como servicio (CaaS), para luego utilizar el framework Hadoop para realizar procesos distribuidos de grandes volúmenes de datos.

Formación de Recursos humanos

El equipo de trabajo está compuesto por los seis docentes-investigadores que figuran en este trabajo y cuatro alumnos.

Se están realizando cuatro tesinas de licenciatura una sobre evaluación de algoritmos de algebra lineal sobre arquitecturas diversas, otra sobre Cloud Computing Privado, otra sobre dispositivos de juegos aplicados a salud y otra sobre SOA aplicada a Cloud. Se espera realizar también una tesis de maestría sobre Metodologías de desarrollo aplicadas a SaaS, otra sobre bases de datos NoSQL y otra sobre algoritmos de Cómputo Intensivo para Big Data y su implementación en Clouds. Además aumentar el número de publicaciones. Por otro lado también se prevé la divulgación de varios temas investigados por medio de cursos de postgrado y actualización o publicaciones de divulgación.

Referencias

- [1] Martin, A., Chavez, S., Murazzo, M., Rodríguez, N., Valenzuela, A. “MongoDB en ambiente Cloud Híbrido con OpenStack”. Wicc2015.
- [2] <http://cassandra.apache.org>
- [3] www.siliconweeks.com
- [4] <https://www.mongodb.com/hadoop-and-mongodb>
- [5] <http://hadoop.apache.org/>
- [6] Murazzo, M., Rodríguez, N., Medel, D., Chavez, S., Martin, A., Valenzuela, F. “Análisis de mejora en la escalabilidad de las infraestructuras de cloud computing”. III Jornadas de Cloud Computing. UNLP (2015).
- [7] Murazzo, M., Tinetti, F., Rodríguez, N. “Despliegue de una Infraestructura Cloud Privada de Código Abierto”. III Jornadas de Cloud Computing UNLP (2015).

ESTUDIO DE INTEGRACIÓN DE MÉTODOS DE DESCUBRIMIENTO DE CONOCIMIENTO EN WEB

Hernán Merlino, Eduardo Diez, Juan Manuel Rodríguez, Santiago Bianco, Ramón García-Martínez

Laboratorio de Investigación y Desarrollo en Arquitecturas Complejas
Grupo Investigación en Sistemas de Información

Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús. Argentina. Tel +54 11 5533 5600 Ext. 5194
hmerlino@gmail.com, rgm1960@yahoo.com

RESUMEN

La extracción de conocimiento a partir de fuentes heterogéneas de información embebida y en volúmenes de datos demasiado grandes, como podría ser la Web, tradicionalmente ha requerido de participación humana en la forma de reglas de extracción o bien de ejemplos de entrenamiento etiquetados de forma manual. Desde hace más de una década se han desarrollado un conjunto de algoritmos como OIE, TEXT RUNNER, WOE-parse, WOE-pos, SRL-Lund, SRLUIUC, ReVerb, TRIPLEX, OLLIE, entre otros, dedicados a la tarea de extracción de conocimiento. En este contexto, este proyecto busca desarrollar un proceso capaz de integrar diversos algoritmos de extracción de conocimiento de forma inteligente, que dado una estructura de información inicial como entrada, que contienen conocimiento embebido, genere un conjunto de piezas de conocimiento (reglas de producción, subgrafos de una red semántica, entre otros).

Palabras clave: Descubrimiento de Conocimiento, Extracción de Conocimiento, Algoritmos de Extracción de Conocimiento.

CONTEXTO

El proyecto: [a] inicia una línea de trabajo en el campo de extracción de conocimiento a partir de fuentes heterogéneas de información embebida dentro del campo de la Informática; [b] articula la Línea de Investigación Prioritaria “3. Desarrollos Informáticos” del Instituto de Economía, Producción y Trabajo, aprobada por Resolución Consejo Superior UNLa N° 113/14, promoviendo la mejora de los sistemas productivos, en particular los utilizados en la producción de tecnologías especiales; y [c] responde a los lineamientos estratégicos de la CADENA DE VALOR DEL SOFTWARE Y SERVICIOS INFORMÁTICOS que establece el Plan Industrial 2020 del Ministerio Industria [MI, 2014], promoviendo la mejora del

sistema productivo de la industria del Software con foco en sistemas software especiales.

INTRODUCCIÓN

El desafío de la extracción de conocimientos comienza a fines de la década de 1970 como es señalado en [Cowie & Lehnert, 1996]. Más tarde en los años 90 la investigación fue alentada y financiada por la Agencia de Proyectos Avanzados de Defensa (DARPA) [Konstantinova, 2014].

Los métodos de extracción de conocimiento comenzaron trabajando en la detección y clasificación de nombres propios, utilizando como entrada fuentes de conocimiento embebido, como son las fuentes de información no estructurada, este tipo de extracción de conocimiento es llamado Reconocimiento de Nombres de Entidades (NER según sus siglas en inglés). En general estos sistemas de extracción de conocimiento buscan nombres de personas, compañías, organizaciones y lugares geográficos [Konstantinova, 2014]. El siguiente paso que dieron los métodos de extracción de conocimiento fue el de resolver correferencias y el de extraer relaciones entre nombres de entidades [Jurafsky & Martin, 2009].

Hacia fines de la década de 2000 los métodos de extracción de conocimiento se habían diversificado y especializado. En [Jurafsky & Martin, 2009] se reconocen distintos tipos de piezas de conocimiento susceptibles de ser extraídas: nombres de entidades, expresiones temporales, valores numéricos, relaciones entre entidades y expresiones previamente identificadas, eventos, entre otras.

La extracción de conocimiento tradicionalmente ha requerido de participación humana en la forma de reglas de extracción o bien de ejemplos de entrenamiento etiquetados de forma manual. En particular para los casos de extracción de relaciones entre entidades, es el usuario quien debe explícitamente, especificar cada relación que le interese, tarea ardua, sobre todo cuando se trabaja con fuentes de información embebida heterogéneas y con volúmenes de datos demasiado grandes, como

podría ser la Web. Debido a ello en general los sistemas de extracción de conocimiento fueron utilizados sobre fuentes de información embebida más bien pequeñas y homogéneas [Banko et al., 2007].

En el año 2007 Michele Banko introduce un nuevo concepto en materia de extracción de conocimiento, al que llama en inglés: Open IE (OIE). Se trata de un paradigma de extracción de conocimiento en donde un sistema informático realiza una sola pasada sobre el total de las fuentes de conocimiento embebido dadas como entrada y extrae un gran conjunto de tuplas relacionales sin requerir ningún tipo de participación humana. En el mismo trabajo Banko presenta un método llamado TEXT RUNNER, el cual es el primer método que trabaja dentro de este nuevo paradigma [Banko et al., 2007].

A partir de este trabajo se propusieron otros métodos de extracción de conocimiento bajo el paradigma que Banko llamó Open IE. En [Wu & Weld, 2010] se propusieron dos métodos WOE-parse y WOE-pos, el primero WOE-parse utiliza un enfoque ligeramente distinto, utiliza un árbol de dependencias, realizando un análisis sintáctico en cada oración, para extraer las relaciones. Y si bien logra un mayor número de extracciones que TEXT RUNNER (1.42 tuplas por oración frente a 0.75) es 30 veces más lento que su predecesor. WOE-pos por el contrario es igual de rápido que TEXT RUNNER y ligeramente mejor (1.05 tuplas extraídas por oración). Si bien WOE-parse y WOE-pos, son métodos de propósito general su base de entrenamiento fue Wikipedia.

En [Mesquita et al., 2010] se presentó un método de extracción de conocimiento llamado SONEX pensado para extraer relaciones de redes sociales y de la blogosfera.

En [Christensen et al., 2011] se presentó un nuevo enfoque bajo las mismas consignas de extracción de conocimiento planteadas por Banko, se buscó utilizar la técnica de etiquetamiento secuencial, basado en la función semántica (en inglés Semantic Role Labeling) para la extracción de relaciones entre entidades. Se crearon dos métodos nuevos SRL-Lund y SRL-UIUC, se los comparó con TEXT RUNNER en dos conjuntos de fuentes de conocimiento embebido, uno pequeño y otro grande. Ambos demostraron ser más precisos que TEXT RUNNER, SRL-Lund obtuvo una precisión de 0.7 y una medida F1 de 0.59, SRL-UIUC obtuvo una precisión de 0.63 y una medida F1 de 0.68 mientras que TEXT RUNNER obtuvo una precisión de 0.55 y una medida F1 de 0.35. Sin embargo, al trabajar con el conjunto más grande de datos de entrada,

TEXT RUNNER demostró tener una ventaja adicional, era 20 veces más rápido que SRL-LUND y 500 veces más rápido que SRL-UIUC.

Ese mismo año, en [Fader et al., 2011] se propone un nuevo método de extracción de conocimiento que logra un área bajo la curva ROC mayor que WOE-parse, WOE-pos y que TEXT RUNNER, se trata de ReVerb. ReVerb fue puesto a prueba utilizando un conjunto de conocimiento embebido que constaba de 500 millones de sentencias web, demostró ser más rápido incluso que TEXT RUNNER. En un subconjunto de 100 000 sentencias se obtuvieron los siguientes tiempos: WOE-parse tardó 11 horas, WOE-pos y TEXT RUNNER tardaron 21 minutos cada uno y ReVerb 16 minutos. La mejora introducida por Fader constó en agregar restricciones a TEXT RUNNER y centrarlo en la extracción de relaciones basadas en verbos.

La restricción de ReVerb no le permite encontrar relaciones basadas en otro tipo de palabras, de categorías gramaticales, que no sean verbos, es por eso que algunos autores han propuesto diversos métodos para extraer relaciones basadas en otro tipo de categorías gramaticales, en particular sustantivos, es el caso de [Schmitz & Soderland, 2012] con OLLIE, de [Yahya et al., 2014] con ReNoun y el de [Mirrezaei et al., 2015] con TRIPLEX. OLLIE fue planteado directamente como una mejora a ReVerb, siendo su objetivo encontrar relaciones basadas no solo en verbos, sino también en sustantivos y adjetivos. Además plantea la posibilidad de hacer un análisis del contexto para encontrar relaciones no explícitas. OLLIE logró obtener 2.7 veces más área sobre la curva ROC que ReVerb y 1.9 veces más área bajo la curva ROC que WOE-parse, además OLLIE encontró 4.4 veces más extracciones correctas que ReVerb y 4.8 veces más que WOE-parse [Schmitz & Soderland, 2012]. El enfoque de TRIPLEX es ligeramente distinto ya que funciona como un complemento a ReVerb o a OLLIE, en el estudio realizado en [Mirrezaei et al., 2015], TRIPLEX por sí solo no logra superar a OLLIE (se comparó utilizando la medida F1 en este caso) y es el uso conjunto de OLLIE más TRIPLEX el que arroja un mejor resultado, aunque no muy lejano al que arroja OLLIE solo.

Para concluir en el trabajo de [Del Corro & Gemulla, 2013] se presenta un nuevo método de extracción de conocimiento llamado ClauseIE (respetando el paradigma propuesto por Banko); en dicho trabajo ClauseIE es comparado contra ReVerb, OLLIE, TEXT RUNNER y WOE utilizando distintas fuentes de conocimiento embebido: 500 oraciones extraídas del conjunto de

datos de prueba utilizado con ReVerb en [Fader et al., 2011], 200 oraciones aleatorias extraídas de Wikipedia y 200 oraciones aleatorias extraídas del New York Times. El resultado en todos los casos fue favorable a ClauseIE quien obtuvo una mejor precisión que los demás métodos.

CONVENCIONES, PREGUNTAS PROBLEMA, HIPÓTESIS Y OBJETIVOS GENERALES Y ESPECÍFICOS

Convenciones

Término: “pieza de conocimiento”

Estructura de información a la que se le puede asignar un significado y que tiene la propiedad de ser automáticamente manipulable en procesos de razonamiento automático. Podría ser una regla de producción o un subgrafo en una red semántica [García-Martínez & Britos, 2004; Gómez et al., 1997]. Dicha estructura representa por lo general relaciones entre conceptos.

Término: “conocimiento embebido”

Conocimiento disponible en estructuras de información entendible por el ser humano pero no por un sistema informático [García-Martínez & Britos, 2004; Gómez et al., 1997]. Un ejemplo de esta clase de conocimiento es el que se encuentra embebido en expresiones del lenguaje natural.

Término: “extracción de conocimiento”

Es el proceso que hace explícito el conocimiento embebido en una estructura de información. Este proceso no requiere especificar las relaciones dentro de las piezas de conocimiento, eventualmente descubre relaciones mientras hace una sola pasada por las estructuras de información que tienen el conocimiento embebido [Banko et al., 2007]. Si se piensa a la extracción de conocimiento como una transformación algebraica podría plantearse:

$$\text{extracción_de_conocimiento}(\text{estructuras_de_información}) = \text{piezas_de_conocimiento}.$$

Preguntas Problema:

¿Es posible crear procesos capaces de integrar diversos métodos de extracción de conocimiento de forma inteligente, teniendo en cuenta las fortalezas y debilidades de cada uno a partir un un conjunto de estructuras de información de dado?

Hipótesis:

Hipótesis I:

Existen diversos y efectivos métodos de extracción de conocimiento en forma de subgrafos. Desde el año 2007 se han propuesto, entre otros: TextRunner [Banko et al., 2007], WOE-parse y WOE-pos [Wu & Weld, 2010], ReVerb [Fader, 2011], OLLIE

[Schmitz & Soderland, 2012], TRIPLEX [Mirrezaei et al., 2015]. Si bien recientemente se han publicado estudios que formulan algunas comparaciones [Del Corro & Gemulla, 2013], no existen estudios comparativos exhaustivos de la calidad de las piezas de conocimiento extraídas por los distintos métodos que refiere la literatura.

Hipótesis II:

Los métodos de extracción de conocimiento están pensados para trabajar con grandes estructuras de información bajo el supuesto de que la calidad de las piezas de conocimiento obtenidas son independientes del dominio. Sin embargo hay indicios experimentales [Schmitz & Soderland, 2012; Mirrezaei et al., 2015] de que esta independencia no es tal [Lopez-Nocera, 2012].

Hipótesis III:

En trabajos de investigación recientes [Del Corro & Gemulla, 2013; Mirrezaei et al. 2015], se plantea la posibilidad de combinar métodos para obtener una sinergia entre los mismos que redunde en una mayor calidad de las piezas de conocimiento obtenidas. Sin embargo no hay estudios comparativos sobre integración de métodos de extracción de conocimiento en general.

Objetivo General:

El objetivo de este proyecto es construir una familia de métodos de extracción de conocimiento tal que dado una estructura de información inicial como entrada, que contienen conocimiento embebido, genere un conjunto de piezas de conocimiento (reglas de producción, subgrafos de una red semántica, entre otros).

Objetivos Específicos:

Objetivo específico vinculado a la Hipótesis I (OE1):

Realizar una comparación entre los distintos métodos de extracción de conocimiento relevados en la literatura, indicando, para diversas estructuras de información, la calidad y los tiempos de ejecución asociados a cada uno.

Objetivo específico vinculado a la Hipótesis II (OE2):

Identificar fortalezas y debilidades en los diversos métodos evaluados en OE1 con el fin de detectar las condiciones particulares bajo las cuales es conveniente utilizar un método por sobre otro.

Objetivo específico vinculado a la Hipótesis III (OE3):

Desarrollar una familia de métodos integrados de extracción de conocimiento que exhiban un mejor

comportamiento que los métodos individuales integrados.

METODOLOGÍA DE TRABAJO

Para construir el conocimiento asociado al presente proyecto de investigación, se seguirá un enfoque de investigación clásico [Riveros y Rosas, 1985; Creswell, 2002] con énfasis en la producción de tecnologías [Sábato y Mackenzie, 1982]; identificando métodos, materiales y abordaje metodológico necesarios para desarrollar el proyecto:

Métodos:

Revisiones Sistemáticas:

Las revisiones sistemáticas [Argimón, 2004] de artículos científicos siguen un método explícito para resumir la información sobre determinado tema o problema. Se diferencia de las revisiones narrativas en que provienen de una pregunta estructurada y de un protocolo previamente realizado.

Prototipado Evolutivo Experimental (Método de la Ingeniería):

El prototipado evolutivo experimental [Basili, 1993] consiste en desarrollar una solución inicial para un determinado problema, generando su refinamiento de manera evolutiva por prueba de aplicación de dicha solución a casos de estudio (problemáticas) de complejidad creciente. El proceso de refinamiento concluye al estabilizarse el prototipo en evolución.

Materiales:

Para el desarrollo de los formalismos y procesos propuestos se utilizarán:

- Formalismos de modelado conceptual usuales en la Ingeniería de Software [Rumbaugh et al., 1999; Jacobson et al., 2013] y en la Ingeniería del Conocimiento [García-Martínez y Britos, 2004].
- Modelos de Proceso usuales en Ingeniería de Software [IEEE, 1997; ANSI/IEEE, 2007; Oktaba et al., 2007].

Abordaje Metodológico:

Para el desarrollo de esta propuesta técnica se han previsto utilizar las siguientes metodologías de investigación y desarrollo:

Para el Objetivo OE1 se propone: (i) realizar una investigación documental exploratoria buscando identificar los métodos de extracción de conocimiento en la literatura reciente; (ii) en base a los resultados del punto precedente identificar las distintas estructuras de información utilizadas para la evaluación de los distintos métodos; (iii) implementar los métodos relevados y ejecutarlos en las estructuras de información características

identificadas; (iv) utilizar la metodología propuesta en [Bronzi et al., 2012] para la cuantificación de la calidad de los distintos métodos.

Para el Objetivo OE2 se propone: (i) realizar una investigación documental exploratoria buscando identificar diferencias en la calidad de las piezas de conocimiento obtenidas por los diversos métodos; (ii) reproducir los experimentos que evidencian las diferencias observadas; (iii) identificar las causas de las diferencias observadas; (iv) identificar que otros métodos son afectados por las mismas causas; (v) proponer nuevos experimentos basados en las características intrínsecas de los métodos de extracción de conocimiento relevados; (vi) identificar nuevas causas que evidencien diferencias entre los algoritmos; (vii) repetir los puntos anteriores hasta que se tenga un conjunto razonable de causas o situaciones que evidencien

Para el Objetivo OE3 se propone: (i) realizar una investigación documental exploratoria buscando identificar que métodos de extracción de conocimiento han sido integrados o utilizados de forma conjunta para mejorar la calidad de las piezas extraídas; (ii) identificar que otros métodos podrían ser integrados para ampliar la calidad o bien la cantidad de las piezas extraídas; (iii) identificar las circunstancias puntuales bajo las cuales la integración de métodos logra extraer más piezas de conocimiento o de mejor calidad que otro grupo de métodos integrados; (iv) desarrollar por el método de prototipado evolutivo un proceso automatizable, que bajo los supuestos anteriores, integre diversos métodos de extracción de conocimiento intentando maximizar la calidad y los tiempos totales según las necesidades del usuario.

RESULTADOS OBTENIDOS/ESPERADOS

El proyecto prevé formular aportaciones conceptuales en el área de métodos de descubrimiento de conocimiento en web. Contar con herramientas de este tipo permite explorar aplicaciones web en metabuscadores y en dispositivos adaptativos de explotación de información.

FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por dos investigadores formados, dos investigadores en formación, y un asesor en metodología de la investigación. En su marco se desarrollan una Tesis de Doctorado en Ciencias Informáticas y una Tesis de Maestría en Tecnología Informática.

FINANCIAMIENTO

Las investigaciones que se proponen en esta comunicación cuentan con financiamiento como Proyecto 80020150200065LA de la Secretaría de Ciencia y Técnica de la Universidad Nacional de Lanús (Argentina).

REFERENCIAS

- ANSI/IEEE, 2007. Draft IEEE Standard for software and system test documentation. ANSI/IEEE Std P829-2007.
- Argimón, J. 2004. Métodos de Investigación Clínica y Epidemiológica. Elsevier España. 84-8174-709-2.
- Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., & Etzioni, O. (2007, January). Open information extraction for the web. In IJCAI (Vol. 7, pp. 2670-2676).
- Basili 1993. The Experimental Paradigm in Software Engineering. En *Experimental Software Engineering Issues: Critical Assessment and Future Directions* (Ed. Rombach, H., Basili, V., Selby, R.). Lecture Notes in Computer Science, Vol. 706. ISBN 978-3-540-57092-9.
- Britos, P. 2008. Procesos de Explotación de Información Basados en Sistemas Inteligentes. Tesis de Doctorado en Ciencias Informáticas. Facultad de Informática. Universidad Nacional de La Plata.
- Bronzi, M., Guo, Z., Mesquita, F., Barbosa, D., & Merialdo, P. (2012, June). Automatic evaluation of relation extraction systems on large-scale. In *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction* (pp. 19-24). Association for Computational Linguistics.
- Christensen, J., Soderland, S., & Etzioni, O. (2011, June). An analysis of open information extraction based on semantic role labeling. In *Proceedings of the sixth international conference on Knowledge capture* (pp. 113-120). ACM.
- Cowie J, Lehnert W (1996) Information extraction. *Communications of the ACM* 39(1):80–91.
- Creswell, J. 2002. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Prentice Hall. ISBN 10: 01-3613-550-1.
- Del Corro, L., & Gemulla, R. (2013, May). ClausIE: clause-based open information extraction. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 355-366). International World Wide Web Conferences Steering Committee.
- Etzioni, O., Cafarella, M., Downey, D., Popescu, A. M., Shaked, T., Soderland, S., ... & Yates, A. (2005). Unsupervised named-entity extraction from the web: An experimental study. *Artificial intelligence*, 165(1), 91-134.
- Fader, A., Soderland, S., & Etzioni, O. (2011, July). Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 1535-1545). Association for Computational Linguistics.
- García-Martínez, R. & Britos, P. V. (2004). *Ingeniería de sistemas expertos*. Nueva Librería. ISBN 987-1104-15-4.
- Gómez, A., Juristo, N., Montes, C., & Pazos, J. (1997). *Ingeniería del conocimiento*. Editorial Centro de Estudios Ramón Areces. ISBN 84-8004-269-9.
- IEEE, 1997. IEEE Standard for Developing Software Life Cycle Processes. IEEE Std 1074-1997 (Revision of IEEE Std 1074-1995; Replaces IEEE Std 1074.1-1995)
- Jurafsky D, Martin JH (2009) *Speech and language processing an introduction to natural language processing, computational linguistics, and speech recognition*, 2nd edn. Prentice-Hall, Inc.
- Konstantinova, N. (2014). Review of Relation Extraction Methods: What Is New Out There?. In *Analysis of Images, Social Networks and Texts* (pp. 15-28). Springer International Publishing.
- Lopez-Nocera, M., Britos, P., Rodriguez, D., Garcia-Martinez, R. 2012. Impacto de la Complejidad del Dominio en las Variaciones del Comportamiento de Procesos de Explotación de Información. *Proceedings IX Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento*. Pág. 55-62. Sello Editorial de la Pontificia Universidad Católica del Perú. ISBN 978-612-4057-85-4.
- Mesquita, F., Merhav, Y., & Barbosa, D. (2010). Extracting information networks from the blogosphere: State-of-the-art and challenges. In *Proceedings of the Fourth AAAI Conference on Weblogs and Social Media (ICWSM), Data Challenge Workshop*.
- Mirrezaei, S. I., Martins, B., & Cruz, I. F. (2015). The Triplex Approach for Recognizing Semantic Relations from Noun Phrases, Appositions, and Adjectives. In *The Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data (Know@LOD) co-located with Extended Semantic Web Conference (ESWC)*, Portoroz, Slovenia.
- Oktaba, H., Garcia, F., Piattini, M., Ruiz, F., Pino, F., Alquicira, C. 2007. Software Process Improvement: The Competisoft Project. *IEEE Computer*, 40(10): 21-28. ISSN 0018-9162.
- Riveros, H. y Rosas, L. 1985. *El Método Científico Aplicado a las Ciencias Experimentales*. Editorial Trillas. México. ISBN 96-8243-893-4.
- Rumbaugh, J., Jacobson, I., Booch, G. 1999. *The Unified Modeling Language, Reference Manual*. Addison Wesley, ISBN-10: 02-0130-998-X.
- Sábato, J. y Mackenzie, M. 1982. *La Producción de Tecnología*. Editorial Nueva Imagen. México. ISBN 968-429-348-8.
- Schmitz, M., Bart, R., Soderland, S., & Etzioni, O. (2012, July). Open language learning for information extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 523-534). Association for Computational Linguistics.
- Wu, F., & Weld, D. S. (2010, July). Open information extraction using Wikipedia. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 118-127). Association for Computational Linguistics.
- Yahya, M., Whang, S. E., Gupta, R., & Halevy, A. (2014, October). Renoun: Fact extraction for nominal attributes. In *Proc. 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar.

Aprendizaje Automático aplicado a Reconocimiento de Patrones en Video y Minería de Datos

Laura Lanzarini¹, Waldo Hasperué¹, César Estrebou¹, Franco Ronchetti^{1,2}, Augusto Villa Monte^{1,2}, Germán Aquino^{1,3}, Facundo Quiroga^{1,2}, María José Basgall^{1,3}, Luis Rojas⁴, Patricia Jimbo Santana⁵, Aurelio Fernandez Bariviera⁶, José Ángel Olivas Varela⁷

¹ Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática - Universidad Nacional de La Plata

² Becario postgrado UNLP ³ Becario postgrado CONICET

⁴ Facultad de Ingeniería. Universidad Nacional de la Patagonia San Juan Bosco

⁵ Dpto. Ciencias de la Computación, Universidad de las Fuerzas Armadas ESPE, Ecuador

⁶ Facultat de Ciències Econòmiques i Empresariales, Universitat Rovira i Virgili Tarragona, España

⁷ Escuela Superior de Informática, Universidad de Castilla-La Mancha, España

{laural, whasperue, cesarest, fronchetti, avillamonte, gaquino, fquiroga, mjbasgall}@lidi.info.unlp.edu.ar
luisf.09@gmail.com, pjimbo@pepsolutions.com, aurelio.fernandez@urv.net, joseangel.olivas@uclm.es

CONTEXTO

Esta presentación corresponde al Subproyecto “Sistemas Inteligentes” perteneciente al proyecto “Cómputo paralelo de altas prestaciones. Fundamentos y evaluación de rendimiento en HPC. Aplicaciones a sistemas inteligentes, simulación y tratamiento de imágenes” (Periodo 2014–2017) del Instituto de Investigación en Informática LIDI.

RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de Sistemas Inteligentes basados en Redes Neuronales Artificiales aplicadas al Reconocimiento de Patrones y a la Minería de Datos.

Con respecto al Reconocimiento de Patrones, el énfasis está puesto en el análisis de videos con el objetivo de identificar acciones humanas que faciliten la interfaz hombre/máquina y en la detección de patrones de movimiento en videos provenientes de cámaras de vigilancia asociados con situaciones de riesgo.

En el área de la Minería de Datos se está trabajando en la generación de un modelo de fácil interpretación a partir de la extracción de reglas de clasificación que permita justificar la toma de decisiones y en desarrollo de nuevas estrategias aplicables a flujos de datos.

Con respecto a Minería de Textos se han desarrollado métodos capaces de extraer las palabras clave de documentos independientemente del lenguaje. Actualmente se están desarrollando estrategias para resumir documentos a través de la extracción de los párrafos más representativos utilizando técnicas de optimización.

Palabras clave: Redes Neuronales, Técnicas de Optimización, Estrategias adaptativas, Reconocimiento de Patrones, Minería de Datos, Minería de Textos.

1. INTRODUCCION

El Instituto de Investigación en Informática LIDI tiene una larga trayectoria en el estudio, investigación y desarrollo de Sistemas Inteligentes basados en distintos tipos de estrategias adaptativas. Los resultados obtenidos han sido medidos en la solución de problemas pertenecientes a distintas áreas. A continuación se detallan los resultados obtenidos durante el último año.

1.1. Reconocimiento de Patrones

En el III LIDI, desde hace varios años se vienen trabajando en procesamiento de señales de audio y video. Como resultado de estas investigaciones se han diseñado e

implementado técnicas originales aplicables al reconocimiento de distintas medidas biométricas así como de gestos dinámicos. Continuando con esta línea, actualmente se están desarrollando los siguientes temas:

Reconocimiento de gestos

La aparición de nuevas tecnologías en sensores y la popularidad de los dispositivos móviles han introducido nuevas posibilidades de interacción hombre-máquina y obtención de información. No obstante, la habilidad de utilizar efectivamente estos datos y modos de comunicación requiere novedosas técnicas de procesamiento, análisis, e inferencia a partir de señales.

En esta línea de investigación, se ha trabajado en la búsqueda de descriptores para acciones humanas dinámicas que permitan, junto con un modelo de clasificación adecuado, un reconocimiento efectivo de las acciones capturadas en videos [1].

Luego de explorar el reconocimiento de gestos y de acciones humanas en general [1,2], en el III-LIDI se ha re-enfocado el objetivo de la investigación en técnicas de reconocimiento de señales temporales complejas para obtener reconocedores de lenguaje de señas robustos.

Se crearon dos bases de datos: la primera, de configuraciones de manos del lenguaje de señas argentino y la segunda de señas dinámicas que utilizan esas configuraciones. Su creación tiene el objetivo de poder realizar experimentos con distintos métodos de reconocimiento y de posibilitar la creación de software que traduzca automáticamente desde el video al lenguaje castellano.

En esta línea de investigación, se publicaron los detalles de la base de datos de configuraciones, junto con un experimento de desempeño inicial que compara métodos tradicionales del estado del arte con un modelo que utiliza descriptores basados en la transformada de Radón y el ProbSom como clasificador [3]. Actualmente, se está trabajando en la clasificación de las señas dinámicas de la segunda base de datos.

Detección de patrones en movimiento en video

Esta línea se ha centrado principalmente en el procesamiento y reconocimiento de patrones en señales de audio e imágenes con el fin de identificar a una persona a través de su rostro o el sonido de su voz. Los resultados obtenidos se encuentran publicados en [4,5,6].

En el último tiempo, se ha comenzado a estudiar la posibilidad de incorporar estos resultados al análisis de videos. Se espera poder contribuir al diseño y desarrollo de nuevas estrategias adaptativas aplicables al análisis la escena. Interesa especialmente investigar aquellas estrategias que permitan reconocer automáticamente la estructura de la escena, las actividades y los patrones de movimiento con el objetivo de cambiar el seguimiento de objetos de bajo nivel por el análisis de eventos de alto nivel y la detección de situaciones anómalas.

Los resultados obtenidos serán utilizados en la detección automática de situaciones de riesgo o amenazas en escenas captadas a través de sistemas de video-vigilancia.

1.2. MINERÍA DE DATOS

Obtención de Reglas de Clasificación

Esta línea de investigación está centrada en la obtención de reglas de clasificación, del tipo IF-THEN, a partir de redes neuronales y técnicas de optimización.

En especial se estudian métodos de clustering y clasificación de patrones que permitan identificar aquellos atributos relevantes para el problema. Dichos atributos serán especialmente considerados en el momento de construir el antecedente de la regla. Reducir la cantidad de atributos a considerar ayuda a obtener un conjunto de reglas sencillas de baja cardinalidad, facilitando de esta forma el criterio que da lugar a la clasificación.

Como área de transferencia tecnológica se ha analizado la situación de las compañías financieras al momento de determinar el riesgo en una operación de otorgamiento de crédito. En estos casos, el proceso de toma de

decisiones habitual consiste en la recopilación de información personal y financiera sobre el prestatario con el objetivo de cuantificar el riesgo que corre la empresa otorgándole el dinero solicitado. Este enfoque no siempre ofrece buenos resultados.

La aplicación del método definido en [7] fue mejorado en [8] al reemplazar la red neuronal SOM por una red neuronal LVQ que por ser supervisada da lugar a una mejor inicialización de la técnica de optimización que determina el conjunto de reglas.

Luego en [9] se utilizó esta nueva estrategia para obtener reglas asociadas a riesgo crediticio demostrando que posee una reducción en el error de Tipo I mejorando de esta forma la identificación de los futuros clientes de la compañía.

Procesamiento de Flujo de Datos

Esta línea de investigación está centrada en el procesamiento de grandes volúmenes de tweets. Los temas abordados en esta línea comprenden el procesamiento del lenguaje natural, la detección de tópicos y el análisis de sentimiento.

Dado el gran volumen de tweets que se genera constantemente se están desarrollando estrategias de procesamiento que involucran el análisis de flujo continuo de tweets otorgando una respuesta rápida la cual se adapta dinámicamente ante la llegada de nuevos tweets.

Se están desarrollando técnicas de procesamiento distribuido en el paradigma MapReduce. Estas técnicas están enfocadas en la tarea de clustering dinámico. El objetivo de las mismas es poder variar los centros dinámicamente agregando y eliminando elementos (palabras) a medida que se detecta la aparición de nuevos tópicos, establecer relaciones entre los tópicos en un momento dado como así también determinar el instante en que determinadas palabras dejan de utilizarse [10].

1.3. MINERIA DE TEXTOS

Extracción de palabras clave en documentos de texto

Esta línea de investigación tiene su eje central en el estudio y aplicación de distintos métodos de representación de documentos así como de distintas técnicas adaptativas aplicables en la resolución de problemas de extracción de palabras clave, tarea de sumo interés ya que permite caracterizar un documento facilitando su búsqueda y clasificación.

En esta línea, se desarrolló un método de identificación de palabras clave a partir de documentos de texto en español utilizando redes neuronales inicialmente descrito en [11] y posteriormente mejorado a través del uso de redes auto-asociativas o autoencoders, los cuales son adecuados en problemas de clasificación muy desbalanceados.

Este método define una representación vectorial para los términos, para luego aplicar un proceso de filtrado gramatical con el fin de remover términos inválidos. Es importante destacar que ésta es la única parte del método que es dependiente del idioma en cuestión.

Una vez obtenida la representación se utiliza un ensemble de autoencoders combinados con redes neuronales convencionales para construir un modelo de clasificación. Dado un término de un documento nuevo los autoencoders determinan para el mismo un error de reconstrucción, que es utilizado para definir un ranking de términos.

Con el objeto de medir la eficacia del método propuesto se utilizaron artículos científicos de WICC y CACIC. Los resultados obtenidos publicados en [12] permiten afirmar que el método propuesto es comparable o incluso superior a Maui, uno de los algoritmos más reconocidos en esta problemática.

Actualmente se continúa el desarrollo del método para mejorar su precisión y ampliar su dominio de aplicación.

Síntesis automática de documentos

En este trabajo de investigación se busca identificar y representar los temas relevantes de un documento. Esto resulta de sumo interés para que a partir de uno o varios textos se pueda comprender la temática general tratada en el mismo, obtener las ideas centrales, categorizarlos automáticamente o comparar su contenido. Anteriormente se desarrolló una estrategia para medir la relevancia de los términos más utilizados en un conjunto de e-mails [13] que permitió mejorar la atención de los alumnos.

Actualmente se está trabajando en la representación de una determinada colección de documentos de texto con el objetivo de identificar las partes relevantes de cada uno de ellos. Para ello se eligió una serie de tesis escritas en LaTeX con temas en común. Se las procesó de manera de tener cada sentencia identificada por su contenido y ubicación dentro del documento, permitiendo que se les calcule un conocido conjunto de métricas. Por otro lado para cada sentencia se estableció la puntuación considerada al momento de construir la síntesis del documento.

A partir de la representación obtenida, se están evaluando distintas técnicas de optimización basadas en cúmulos de partículas con el fin de identificar el aporte de cada métrica en la puntuación propuesta. Esto permite encontrar la combinación adecuada de métricas que dan lugar al resumen esperado de cada documento.

2. TEMAS DE INVESTIGACIÓN Y DESARROLLO

- Estudio de técnicas de optimización poblaciones y redes neuronales artificiales para la obtención de reglas de tipo IF-THEN.
- Clasificación de configuraciones de manos para el lenguaje de señas.
- Desarrollo de representaciones para el reconocimiento de señas.
- Métodos estructurados y no estructurados

a la representación de documentos.

- Problemas de clasificación con desbalance de clase severo.
- Síntesis de documentos de texto a partir de la selección de las partes relevantes obteniendo la combinación de un conjunto de métricas.
- Implementación de técnicas para el tratamiento de flujos continuos de tweets para la obtención de respuestas dinámicas.
- Implementación de técnicas en el paradigma de MapReduce para clustering dinámico.

3. RESULTADOS OBTENIDOS

- Desarrollo de un método de obtención de reglas de clasificación con énfasis en la reducción de la complejidad del modelo aplicable a riesgo crediticio.
- Elaboración de una base de datos de configuraciones de manos del Lenguaje de Señas Argentino, y evaluación de su desempeño con varios métodos.
- Elaboración de una base de datos de señas dinámicas del Lenguaje de Señas Argentino.
- Desarrollo de una representación de términos que junto con un modelo de clasificación permite identificar palabras clave en un documento.
- Implementación de una técnica para la tarea de clustering dinámico en el paradigma MapReduce.
- Análisis de sentimiento de tweets relacionados con los tópicos del cambio climático en relación a la cumbre de París (COP21) realizada en noviembre-diciembre de 2015. Esta tarea se está realizando conjuntamente con integrantes del departamento de Educación de la Universidad de Stanford.

4. FORMACIÓN DE RECURSOS HUMANOS

Dentro de los temas involucrados en esta línea de investigación se han finalizado 2 tesis de doctorado, 2 tesis de maestría y 3 tesis de especialista y numerosas tesinas de grado de Licenciatura.

Actualmente se están desarrollando 6 tesis de doctorado, 1 de maestría y 2 tesinas de grado de Licenciatura. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

5. REFERENCIAS

- [1] Ronchetti F., Quiroga F., Lanzarini L., Estrebou C. *Distribution of action movements (DAM): a descriptor for human action recognition*. Frontiers of Computer Science. Springer - Higher Education Press. Pp 1-10. ISSN 2095-2228. Septiembre 2015
- [2] Quiroga F., Corbalán L. *A novel competitive neural classifier for gesture recognition with small training sets*. XVIII Congreso Argentino de Ciencias de la Computación. 2013. pp. 140-149.
- [3] Ronchetti F., Quiroga F., Estrebou C., Lanzarini L. *Clasificación de configuraciones de manos del Lenguaje de Señas Argentino con ProbSOM*. XXI Congreso Argentino de Ciencias de la Computación. CACIC 2015.
- [4] Maulini J., Lanzarini L. *Face Recognition using SIFT descriptors and Binary PSO with velocity control*. Computer Science & Technology Series – XVII Argentine Congress of Computer Science Selected Papers. ISBN 978-950-34-0885-8. EDULP, 2012. pp. 43-53.
- [5] Albanesi B., Funes N., Chichizola F., Lanzarini L. *Reconocimiento de Objetos en Video utilizando SIFT paralelo*. XVI Congreso Argentino de Ciencias de la Computación, CACIC 2010.
- [6] Estrebou C., Lanzarini L., Hasperué W. *Voice recognition based on probabilistic SOM*. Conferencia Latinoamericana de Informática. CLEI 2010.
- [7] Lanzarini, L., Villa Monte, A., Ronchetti, F.: SOM+PSO. *A Novel Method to Obtain Classification Rules*. Journal of Computer Science & Technology (JCS&T), Vol. 15, No 1, pp. 15-22. ISSN 1666-6038. Abril 2015.
- [8] Lanzarini L., Villa Monte A., Aquino G., De Giusti A. *Obtaining classification rules using lvqPSO*. Advances in Swarm and Computational Intelligence. Lecture Notes in Computer Science. Vol 6433, pp. 183-193. ISSN 0302-9743. Springer-Verlag Berlin Heidelberg. Junio 2015.
- [9] Lanzarini L., Villa Monte A., Fernandez Bariviera A., Jimbo Santana P. *Obtaining Classification Rules Using LVQ+PSO: an application to Credit Risk*. Scientific Methods for the Treatment of Uncertainty in Social Sciences. Advances in Intelligent Systems and Computing. Springer-Verlag Berlin Heidelberg. vol. 377. Pp 383-391. ISSN 2194-5357. 2015.
- [10] Basgall M. J., Aquino G., Lanzarini L., Naiouf M. "Un Enfoque Dinámico para la Detección de Relaciones entre Tópicos en Textos provenientes de Redes Sociales". III Jornadas de Cloud Computing y Big Data (JCC&BD). Facultad de Informática. UNLP. 29 de junio al 3 de Julio de 2015.
- [11] Aquino, G., Hasperué, W., Estrebou, C., Lanzarini, L. *A Novel Language-Independent Keyword Extraction Method*. Publicado en el Libro *Computer Science & Technology Series – XIX Argentine Congress of Computer Science - Selected Papers*. Editores: Jorge Raúl Finocchietto, Patricia Pesado. EDULP, 2014. Pags.221-232.
- [12] Aquino G., Lanzarini L. *Keyword Identification in Spanish Documents using Neural Networks*. XXI Congreso Argentino de Ciencias de la Computación. CACIC 2015.
- [13] Villa Monte, A. Estrebou, C., Lanzarini, L. *E-mail processing using data mining techniques*. Computer Science & Technology Series – XVI Argentine Congress of Computer Science Selected Papers, ISBN 978-950-34-0757-8. EDULP, 2011. pp. 109-120.

Aplicaciones del aprendizaje automático en psicología y educación

Marcelo Errecalde, Leticia Cagnina,^{*} Edgardo Ferretti, Irma Pianucci,
Ana María Jofré, Maximiliano Sapino,^{**} Luciana Mariñelarena-Dondena^{***}

Laboratorio de Investigación y Desarrollo en Inteligencia Computacional

Departamento de Informática, Universidad Nacional de San Luis

Ejército de los Andes 950 - (D5700HHW) San Luis - Argentina

e-mails de contacto: {merreca,lcagnina,ferretti}@unsl.edu.ar

Resumen

Este artículo describe, brevemente, las tareas de investigación y desarrollo que se llevan a cabo en la línea de investigación “Aplicaciones” en el marco del proyecto “Aprendizaje automático y toma de decisiones en sistemas inteligentes para la Web”. A diferencia de otras líneas de este proyecto, especializadas en aspectos computacionales del Aprendizaje Automático, la Minería de Datos y la Inteligencia Artificial en general, en esta línea participan investigadores de otras áreas como la psicología y la educación. El motivo de esta conformación de la línea, tiene que ver con la naturaleza interdisciplinaria de las aplicaciones abordadas, lo cual requiere conocimiento especializado en estos campos, tanto para la identificación de la información relevante a utilizar, el etiquetado de los datos, la evaluación de tests y cuestionarios y la interpretación de los resultados. Las temáticas que se abordan, incluyen la explotación de información educativa, la explotación de información en estudios psicológicos, la predicción de rasgos de personalidad, emociones, nivel de bienestar y toma de decisiones, la identificación de estilos de aprendizaje y la modelización del estudiante en general, entre otras.

^{*}Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

^{**}Bechario de Doctorado (CONICET)

^{***}Bechario de post-Doctorado (CONICET)

Palabras clave: Minería de Datos Educativa, Análisis Dirigido por los Datos en los Medios Sociales, Ciencia de los Datos, Aprendizaje Automático

Contexto

La línea de investigación “Aplicaciones” es una de las tres líneas del proyecto titulado “Aprendizaje automático y toma de decisiones en sistemas inteligentes para la Web”, un nuevo proyecto que será presentado este año como continuación del Proyecto de Investigación Consolidado (PROICO) titulado “Herramientas y mecanismos para la toma de decisiones en agentes inteligentes artificiales”. Este último proyecto, aprobado por evaluadores externos a la UNSL, se desarrolla en el *Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)* de la UNSL y ha sido financiado en forma directa por la UNSL (PROICO 30312) y en forma indirecta por: a) el Programa de Incentivos (22/F237), b) la Comisión Europea de Investigación e Innovación, a través del programa Marie Curie Actions: FP7 People 2010 IRSES, c) el CONICET, a través de un investigador asistente y becas: dos de Doctorado y dos de Post-Doctorado asignadas a integrantes del proyecto y d) el Consejo Nacional de Ciencia y Tecnología (CONACYT-México) y otros organismos científicos del gobierno me-

xicano, en los que distintos integrantes han participado en tres proyectos de investigación como colaboradores externos.

Este proyecto posee además otras dos líneas de investigación denominadas: “Toma de decisiones y aprendizaje automático” y “Minería de textos y de la Web”; la primera dedicada al desarrollo de modelos formales y mecanismos para la toma de decisiones y aprendizaje en agentes artificiales inteligentes y la segunda vinculada a la ingeniería del lenguaje natural, incluyendo temas como el Procesamiento del Lenguaje Natural (PLN) y la Lingüística Computacional, la Minería de Textos y de la Web y la recuperación de información de la Web. Es claro en este contexto, que muchos problemas y aplicaciones intersectan los alcances de más de una de las líneas de este proyecto, lo cual involucra un trabajo integrado y coordinado permanente a los fines de optimizar los recursos disponibles para la obtención de los objetivos propuestos.

Introducción

Las ciencias sociales han ingresado en la era de la *ciencia de los datos* (*data science*) ya que ahora es posible analizar grandes volúmenes de información que se han hecho disponible a través de los medios sociales (*social media*), plataformas de *e-learning* y otros sistemas accesibles via Web. Esta oportunidad conlleva importantes desafíos, ya que es necesario realizar *investigaciones interdisciplinarias*, por ejemplo entre la computación y la psicología [15] o la computación y la educación [17].

Estas investigaciones interdisciplinarias requieren, por una parte, familiarizar a los especialistas que no provienen del ámbito de las ciencias de la computación, con diversas prácticas metodológicas y terminología que, si bien son usuales en áreas como el aprendizaje automático y la minería de datos, todavía no están tan difundidas en áreas como la psicología y la educación. Por otra parte, los especialistas informáticos en esta nueva “ciencia de los datos”, requieren para la aplicación de sus técnicas de análisis y ex-

plotación de información, de conocimientos muy específicos del área particular abordada, tanto para la definición de las “features” de los objetos a ser analizados y de su “etiquetado”, como de la interpretación de los resultados obtenidos de dicho análisis.

En este contexto, esta línea de investigación está formada por integrantes provenientes no sólo del área del aprendizaje automático y los sistemas inteligentes, sino también por especialistas y becarios especializados en el área de la psicología y la educación, entre otros. El objetivo es proveer un ámbito favorable para el trabajo interdisciplinar, de manera tal de hacer fácilmente disponible las herramientas de explotación de información a las demás disciplinas, y recibir por otra parte, asesoramiento específico en aquellos problemas de disciplinas no vinculadas directamente a la computación.

Es claro en este caso, que esta línea se ve nutrida por los desarrollos computacionales, que se realizan en las otras dos líneas de investigación del proyecto. Sin embargo, y como se verá en la siguiente sección, varios de los problemas que se enfrentan al desarrollar aplicaciones concretas del mundo real, no son correctamente solucionados por los modelos y técnicas actuales provistas por el aprendizaje automático, la minería de datos y la inteligencia artificial en general. Es así, que varios de los problemas enfrentados en estos casos, actuarán como disparadores para nuevos desarrollos teóricos y modelos que serán abordados en las restantes líneas del proyecto.

En la siguiente sección, se describen brevemente algunas de las principales aplicaciones que se abordan en esta línea de investigación, describiéndose además los trabajos en desarrollo y a futuro.

Líneas de Investigación y Desarrollo

En términos generales, esta línea de investigación se enfoca en distintos tipos de aplicaciones que hemos organizado en 2 grandes grupos: 1) explotación de información en psi-

cología y 2) explotación de información en educación. Cada una de ellas se detalla brevemente a continuación.

Explotación de información en psicología

Las ciencias sociales tienen hoy disponible en los medios sociales de la Web, grandes volúmenes de información que pueden ser analizados mediante técnicas de la minería de datos (o *data mining*). Esta oportunidad, conlleva el desafío de realizar investigaciones interdisciplinarias, por ejemplo entre la psicología y la informática [15].

En este contexto, James Pennebaker sostiene que el lenguaje natural que las personas utilizan cotidianamente refleja su personalidad, su situación social y las relaciones interpersonales que entablan con los demás. Incluso afirma que el lenguaje puede servir para evaluar la salud física y mental de los sujetos [18].

En esa dirección, Schwartz y Ungar [16] plantean que tradicionalmente los/as psicólogos/as han evaluado los pensamientos, los sentimientos y los rasgos de personalidad mediante cuestionarios administrados a muestras relativamente pequeñas de participantes. En contraposición, ponen de relieve las nuevas alternativas para la evaluación psicológica que brindan el análisis de contenido conducido por los datos (*data-driven content analyses*) o el enfoque del vocabulario abierto (*open vocabulary approach*), si se utilizan los grandes volúmenes de información disponibles en los medios sociales como Facebook y Twitter.

Al mismo tiempo, otros estudios han demostrado que el análisis del lenguaje disponible en los medios sociales es extraordinariamente útil para realizar estudios epidemiológicos a gran escala o para identificar las características psicológicas que prevalecen en diferentes regiones geográficas, por ejemplo aquellas vinculadas con el bienestar. Este método de evaluación es mucho más rápido y menos costoso que las tradicionales encuestas realizadas por las agencias del gobierno [14].

Como eje de este tipo de estudios, se han

utilizado técnicas clásicas del aprendizaje automático, la minería de datos y de textos y el procesamiento del lenguaje natural. En general, estas investigaciones han abordado tanto *tareas predictivas*, como la clasificación y la regresión, como *descriptivas*, como el agrupamiento (*clustering*) y el análisis de asociaciones y correlaciones. En el primer caso, se construye un *modelo predictivo* (clasificador o regresor) mediante un proceso de aprendizaje automático inductivo de tipo supervisado, que toma como datos de entrenamiento un conjunto de objetos (documentos, usuarios, etc) etiquetados. Este modelo, estima la variable dependiente/salida, usualmente llamada *etiqueta* o *resultado*, en función de determinados *predictores* o *características* (*features*) derivados de los objetos analizados (variables independientes). El modelo predictivo debe ser capaz de clasificar correctamente otros datos que nunca fueron presentados en la etapa de entrenamiento (capacidad de generalización). En el segundo caso, en cambio, se obtienen *modelos descriptivos* que hacen explícitos grupos, patrones y correlaciones entre las variables involucradas que mejoran nuestra comprensión y entendimiento de los datos analizados (*insight*).

En este contexto, en esta línea de investigación nos proponemos el diseño de métodos de evaluación de los rasgos de personalidad, la toma de decisiones, las emociones y el bienestar psicológico para población argentina utilizando técnicas de aprendizaje automático y minería de textos. Además, se evaluará la potencialidad del uso de información multimodal (imágenes, grabaciones de audio, video clips, etc.) disponible en los medios sociales (*social media*) como Facebook y Twitter, mediante herramientas de explotación de información en la Web, para el diseño de métodos de evaluación psicológica. Para llevar a cabo un estudio de esta naturaleza, esta línea cuenta con un becario de Doctorado y una becaria de Post-Doctorado del CONICET, ambos del área de psicología, quienes estarán a cargo de interactuar con los participantes de los experimentos, a los que se le solicitará la redacción de diversos textos dirigidos, y con los que se utilizarán dis-

tintos instrumentos de evaluación psicológica específicos para cada uno de los aspectos psicológicos abordados.

Explotación de información en educación

Este área de estudio, que surge de la aplicación de técnicas de aprendizaje automático y minería de datos en el ámbito educativo, ha sido referenciado de diversas maneras, popularizándose últimamente por expresiones en inglés tales como *Educational Data Mining* (EDM) [12] y *Learning Analytics* [7]. Las motivaciones del uso de este tipo de técnicas en el ámbito educativo son claras: las particularidades de los estudiantes y de su interacción con las distintas tecnologías que soportan el proceso de aprendizaje, pueden hoy en día ser capturadas y analizadas, lográndose de esta manera un nivel de adaptación y personalización de los sistemas vinculados al proceso de enseñanza aprendizaje, que nunca antes había sido logrado. Es así como podemos ver sistemas que se adaptan a los estudiantes de acuerdo a sus estilos de aprendizaje [11], que recomiendan contenidos de acuerdo a la dificultad/facilidad que ha tenido un estudiante en la resolución de los distintos tipos de ejercicios [4] o que predicen la deserción o reprobación de un estudiante en un curso on-line de acuerdo a su desempeño y/o comportamiento previo [2].

Las técnicas de minería de datos pueden aplicarse sobre datos surgidos de los procesos de enseñanza-aprendizaje, tanto en sus variantes *presenciales* como a *distancia*. No obstante esto, su utilización en cursos dictados a distancia o B-learning (Blend Learning) y soportados por *Entornos Virtuales de Enseñanza - Aprendizaje* parece recibir la mayor atención en estos días, con ejemplos concretos llevados a cabo en la plataforma de aprendizaje Moodle [8], o bien en cursos en línea masivos y abiertos (MOOCs por sus siglas en inglés) como por ejemplo los provistos por Coursera [3].

Al igual que en las aplicaciones para psicología, la explotación de información educativa puede ser dividida en aquellas que se

enfocan en tareas *predictivas* como clasificación [13] y regresión [10], o descriptivas como clustering [1] y análisis de reglas de asociación [9]. Como se ha observado en distintos estudios y surveys, si bien en los comienzos se dio preferencia a los modelos predictivos, cada día gana más interés la resolución de problemas que involucran algún tipo de predicción. Esto ha quedado de manifiesto en la que quizás sea una de las principales competencias en el área de extracción de conocimiento como lo es la *Knowledge Data Discovery (KDD) Cup*, que en su versión del año 2010¹ se enfocó en la predicción de la calificación numérica que un estudiante tendrá en un ejercicio particular (regresión), mientras que en 2015², se abordó la predicción de abandonos (dropouts) en un MOOC (clasificación).

En este contexto, en esta línea de investigación nos proponemos la aplicación de técnicas de aprendizaje automático y minería de datos en plataformas tipo Moodle y con información provista por MOOCs del estilo Coursera. Si bien este es un área que hemos comenzado a estudiar muy recientemente, estamos especialmente interesados en abordar aquellos aspectos temporales, que influyen en la modelización del conocimiento que el estudiante va adquiriendo a lo largo del tiempo y que puede impactar en su desempeño. Estos aspectos temporales, también deberían ser considerados a la hora de determinar en la forma más anticipada posible, cuándo un estudiante tiene altas chances de abandonar un curso. Estos aspectos, que pueden ser caracterizados en el contexto de la *categorización anticipada*, han comenzado a ser estudiadas en nuestro grupo en colaboración con investigadores de México [5]. Por otra parte, también se preve la incorporación de este tipo de técnicas en sistemas recomendadores, tarea que será llevado a cabo con la línea de toma de decisiones y aprendizaje en agentes inteligentes artificiales. Además, si bien existe una tendencia en la investigación vinculada a determinar los estilos de aprendizaje,

¹KDD Cup 2010: Educational Data Mining Challenge

²KDD Cup 2015: Predicting dropouts in MOOC. <http://kddcup2015.com/>

nuestra idea es combinar estas propuestas, con aquellas que también consideran aspectos psicológicos, como ya ha sido planteado en [6].

Al momento, para llevar a cabo estos estudios, esta línea cuenta con dos docentes especializadas en informática educativa, las cuales comenzarán en 2016 sus estudios de postgrado y un trabajo final de Licenciatura trabajando en estos temas.

Formación de Recursos Humanos

2 becarios de CONICET trabajando en explotación de información aplicada a psicología:

- Uno de ellos como parte de su tesis de Doctorado a iniciarse en 2016.
- Otro, como parte de sus tareas de Postdoctorado iniciadas en 2015.

Trabajos de tesis vinculados con explotación de información en educación:

- 1 tesis de Licenciatura a iniciarse en 2016.
- 1 tesis de Maestría a iniciarse en 2016.

Referencias

- [1] Profiles of more and less successful {L2} learners: A cluster analysis study. *Learning and Individual Differences*, 22(4):463 – 472, 2012.
- [2] G. Balakrishnan. Predicting student retention in massive open online courses using hidden markov models. Master's thesis, EECS Department, University of California, Berkeley, 2013.
- [3] D. S. Chaplot, E. Rhim, and J. Kim. Predicting student attrition in moocs using sentiment analysis and neural networks. In *Proceedings of the Workshops at the 17th International Conference on Artificial Intelligence in Education (AIED 2015)*, volume 3, 2015.
- [4] A. Corbi and D. Burgos. Implementation of the recommendation model lime in cognitive and visual interactive tutors from pslc. *IEEE Latin America Transactions*, 13(2):516–522, 2015.
- [5] H. J. Escalante, M. Montes-y-Gómez, L. V. Pineda, and M. L. Errecalde. Early text classification: a naive solution. *CoRR*, abs/1509.06053, 2015.
- [6] T. B. Fariba. Academic performance of virtual students based on their personality traits, learning styles and psychological well being: A prediction. *Procedia - Social and Behavioral Sciences*, 84:112 – 116, 2013. The 3rd World Conference on Psychology, Counseling and Guidance, WCPCG-2012.
- [7] R. Ferguson. Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6):304–317, 2012.
- [8] S. Hassan and A. E. F. Hegazy. A model recommends best machine learning algorithm to classify learners based on their interactivity with moodle. In *Computing Technology and Information Management (ICCTIM), 2015 Second International Conference on*, pages 49–54, 2015.
- [9] M. Holzhüter, D. Frosch-Wilke, and U. Klein. *Intelligent and Adaptive Educational-Learning Systems*, chapter Exploiting Learner Models Using Data Mining for E-Learning: A Rule Based Approach, pages 77–105. Springer Berlin Heidelberg, 2012.
- [10] L. Macfadyen and P. Sorenson. Using lims (the learner interaction monitoring system) to track online learner engagement and evaluate course design. In *Proc. of the 3rd Int. Conf. on Educational Data Mining*, pages 301–302, 2010.
- [11] E. Özpolat and G. B. Akar. Automatic detection of learning styles for an e-learning system. *Computers & Education*, 53(2), 2009.
- [12] A. Peña-Ayala. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, 41(4, Part 1):1432 – 1462, 2014.
- [13] P. Ping-Feng, L. Yi-Jia, and W. Yu-Min. Analyzing academic achievement of junior high school students by an improved rough set model. *Computers & Education*, 54(4), 2010.
- [14] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, R. E. Lucas, M. Agrawal, G. J. Park, S. K. Lakshmikanth, S. Jha, M. Seligman, and L. H. Ungar. Characterizing geographic variation in well-being using tweets. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2013.
- [15] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stilwell, M. Seligman, and L. H. Ungar. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE*, 8(9), 2013.
- [16] H. A. Schwartz and L. H. Ungar. Data-driven content analysis of social media - a systematic overview of automated methods. *The ANNALS of the American Academy of Political and Social Science*, 659(1):78–94, 2015.
- [17] S. B. Shum and R. Ferguson. Social learning analytics. *Journal of Educational Technology and Society*, 15(3):3–26, 2012. Special Issue on Learning & Knowledge Analytics.
- [18] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54, 2010.

Aportes al Modelo de Bases de Datos Métricas

Jorge Arroyuelo, Susana Esquivel, Alejandro Grosso, Verónica Ludueña, Cintia Martínez, Nora Reyes
Dpto. de Informática, Fac. de Cs. Físico-Matemáticas y Naturales, Universidad Nacional de San Luis
{bjarroyu, esquivel, agrosso, vlud, nreyes}@unsl.edu.ar, cintiavmartinez@hotmail.com

Edgar Chávez

Centro de Investigación Científica y de Educación Superior de Ensenada, México

elchavez@cicese.mx

Karina Figueroa

Fac. de Cs. Físico-Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo, México

karina@computo.fismat.umich.mx

Gonzalo Navarro

Dpto. de Cs. de la Computación, Universidad de Chile

gnavarro@dcc.uchile.cl

Manuel Hoffhein, Rodrigo Paredes

Dpto. de Cs. de la Computación, Fac. de Ingeniería, Universidad de Talca, Chile

mahahein3@gmail.com, raparede@utalca.cl

Resumen

La computación se ha vuelto indispensable en cualquier ámbito de la vida moderna: ciencias, arte, educación, finanzas, diversión, etc., por lo que se hizo prioritario el desarrollo de aplicaciones capaces de manipular casi cualquier tipo de datos. Para lograr un alcance masivo, muchas de estas aplicaciones son cada vez más intuitivas; por ejemplo, es común ingresar una imagen o un trozo de canción a un buscador y esperar que éste muestre imágenes o canciones parecidas a la provista.

Claramente, para lograr la manipulación eficiente de datos como imágenes, audio, video, secuencias de ADN, texto, huellas digitales, etc., es necesario utilizar depósitos especializados y técnicas de búsquedas no exactas sobre ellos, ya que las soluciones tradicionales no permiten hacer frente a tales requerimientos. Las *Bases de Datos Métricas* son uno de los modelos generales en los cuales se pueden utilizar estructuras de datos especializadas que contemplen estos aspectos. Además de proveer una respuesta rápida y adecuada, será necesario un eficiente uso del espacio disponible, y si se consideran bases de datos masivas, dichas estructuras en particular serán *estructuras de datos con I/O eficiente*.

Otro aspecto importante son los lenguajes de consulta, necesarios para la manipulación de una base de datos, que no siempre poseen el poder expresivo necesario para reflejar las consultas consideradas de interés en este modelo. Así, nuestra investigación pretende contribuir a la consolidación de este nuevo modelo de bases de datos.

Palabras Claves: bases de datos no convencionales, lenguajes de consulta, índices, expresividad.

Contexto

El Proyecto Consolidado 330314 (Código 22/F414 en el Programa de Incentivos a la Investigación), *Tecnologías Avanzadas de Bases de Datos* de la Universidad Nacional de San Luis, y en particular la línea *Bases de Datos no Convencionales*, constituyen el marco en el cual se desarrolla el presente trabajo. Éste se centra en la investigación de aspectos relacionados con la administración de bases de datos capaces de manipular todo tipo de datos. Esto incluye la expresividad de los lenguajes de consulta, los operadores necesarios para responder preguntas de interés, y el análisis de aspectos teóricos, empíricos y aplicativos de los mismos; contribuyendo así a distintos campos de aplicación: sistemas de información geográfica, computación móvil, robótica, visión artificial, motores de búsqueda en internet, diseño asistido por computadora, etc.

La colaboración de nuestros integrantes en investigaciones nacionales e internacionales permite contemplar nuevas perspectivas en nuestros estudios. Se mantiene cooperación con: Universidad de Chile, Universidad de Talca (Chile), Universidad Michoacana de San Nicolás de Hidalgo (México), Centro de Investigación Científica y de Educación Superior de Ensenada (México), Universidad de Massey (Nueva Zelanda) y Universidad Nacional de La Matanza.

Introducción

En la actualidad la computación ha alcanzado todos los ámbitos de la vida moderna, esto ha provocado el desarrollo de aplicaciones capaces de adaptarse tanto a estos nuevos entornos como a los diversos usuarios de las mismas. Esto implica el desarrollo de bases de datos capaces de administrar todo tipo de datos y responder consultas sobre los mismos de una manera totalmente diferente a la tradicional, muchas veces más intuitiva. Algunos ejemplos de aplicaciones: comparación de huellas digitales, bases de datos médicas, reconocimiento de voz, reconocimiento facial, reconocimiento de imágenes, recuperación de texto, biología computacional, minería de datos, clasificación y aprendizaje automático, etc.

Todas estas aplicaciones tienen características comunes, que pueden englobarse en el modelo de *espacios métricos*. Formalmente un espacio métrico consiste de un universo de objetos \mathbb{U} y una función de distancia definida entre ellos $d : \mathbb{U} \times \mathbb{U} \mapsto \mathbb{R}^+$ que mide la disimilitud entre los objetos. En este escenario las búsquedas exactas carecen de sentido, si se introduce un trozo de melodía en un buscador, se espera obtener aquellas que sean similares a ésta. Entonces es importante la elección de este modelo por las *búsquedas por similitud*, más naturales sobre estos tipos de datos.

Para evitar la examinación secuencial de los datos al responder a este tipo de búsquedas, se utilizan los *Métodos de Acceso Métricos* (MAMs). La mayoría de estos métodos no admiten dinamismo, no están diseñados para operaciones de búsqueda complejas, ni para soportar conjuntos masivos de datos, esto permite analizar distintas maneras de optimizarlos. El trabajo con bases de datos masivas, o con aquellas que almacenan objetos muy grandes, da lugar también a líneas de investigación que, considerando el cambio del modelo de costo utilizado, diseñan MAMs más eficientes (en espacio, en I/O, etc.) para memorias jerárquicas. Otra área de investigación explorada es la expresividad de los lenguajes de consulta utilizados, tratando de incrementar la misma para expresar consultas más precisas y caracterizar la clase de consultas computables.

Líneas de Investigación y Desarrollo

Bases de Datos Métricas

Los espacios métricos generales sirven como modelo para las bases de datos que manipulan datos no convencionales (imágenes, videos, texto libre, secuencias de ADN, audio, etc.). En este modelo la

complejidad se mide como el número de cálculos de distancias realizados al crear el índice, o al realizar búsquedas, dado el costo que implica su cálculo. Por ello, y dado que los MAMs son necesarios al momento de responder las diversas consultas a una base de datos, se analizan aquellos que han mostrado buen desempeño en las búsquedas. El objetivo es lograr optimizarlos reduciendo su complejidad y considerando, cuando sea necesario, la jerarquía de memorias. En general, dada una base de datos $X \subseteq \mathbb{U}$ y una consulta $q \in \mathbb{U}$ las consultas son de dos tipos: por *rango* o de *k-vecinos más cercanos*, aunque existen otras operaciones de interés [15].

Métodos de Acceso Métricos

A partir del *Árbol de Aproximación Espacial* [9], un índice que mostró un muy buen desempeño en espacios de mediana a alta dimensión, pero totalmente estático, se desarrolló uno de los pocos índices completamente dinámicos: el *Árbol de Aproximación Espacial Dinámico (DSAT)* [10] que permite realizar inserciones y eliminaciones, conservando su buen desempeño en las búsquedas.

Si en una base de datos métrica, ya sea por ser masiva o porque sus objetos son muy grandes, el índice no cabe en memoria principal, entonces surge la necesidad de hacer uso de la memoria secundaria. Esto requiere diseñar índices especialmente para memoria secundaria. Así, en [11] se presentaron versiones preliminares del *DSAT* (*DSAT+* y *DSAT**), que sólo admiten inserciones y búsquedas eficientes. Sin embargo, numerosas aplicaciones necesitan total dinamismo; es decir, que también puedan realizarse eliminaciones. Así, se ha diseñado un nuevo índice dinámico para memoria secundaria, basado en la *Lista de Clusters* [3], que tiene buen desempeño en espacios de alta dimensión, es completamente dinámico, con buena ocupación de página y sus operaciones son eficientes tanto en cálculos de distancia como en operaciones de I/O [12]. Sin embargo, las búsquedas en este índice deben recorrer completamente la lista de centros de los clusters, lo cual produce costos significativos. Así, combinando esta nueva estructura con lo bueno del *DSAT*, se ha diseñado un nuevo índice que mantiene los centros de los clusters en un *DSAT* y los clusters mismos en memoria secundaria. De esta manera, se han mejorado los costos de las operaciones en cuanto a cálculos de distancia, manteniendo los bajos costos de acceso a disco que se tenían.

Sin embargo, existen otras maneras posibles de lograr un índice totalmente dinámico a partir de la

Lista de Clusters. Por ello, se ha diseñado otro índice que, combinando *algoritmos de pivotes* [3] con “clusters” cuyo tamaño depende del tamaño de página de disco, logra operaciones eficientes.

En algunos casos, aunque la estructura sea eficiente, con el fin de lograr una respuesta más rápida, se intercambia precisión por velocidad en la respuesta. Es decir, se admite que ante una consulta se devuelvan sólo algunos objetos relevantes, siempre que dicha respuesta se encuentre disponible mucho más rápido. Estos tipos de búsquedas se denominan *aproximadas*. Un algoritmo muy eficiente para este tipo de consultas es el llamado *algoritmo basado en permutaciones* [2]. Por lo tanto, se está diseñando un nuevo índice que combine las ideas de [12], pero que agrupe por distancia entre las permutaciones de los objetos, en lugar de por distancia entre objetos. Esto permitiría obtener un índice al que se le pueda indicar el número máximo de cálculos de distancia y/o el número máximo de operaciones de I/O, que se está dispuesto a utilizar, para obtener una respuesta rápida, aunque menos precisa.

Para analizar cuán buenos son los agrupamientos que logran estas estructuras, se pueden utilizar estrategias de optimización basadas en heurísticas bioinspiradas, que sirven para detección de clusters.

Búsqueda Aproximada de los *All-k-NN*

Algunas de las aplicaciones incluídas en el modelo de *espacios métricos* son la clasificación y aprendizaje automático: un nuevo elemento debe ser clasificado de acuerdo a sus vecinos más cercanos; la cuantificación y compresión de imágenes: sólo algunos vectores pueden ser representados y aquellos que no pueden serlo, deben ser codificados como su punto representable más cercano; la predicción de funciones: se desea buscar el comportamiento más similar de una función en el pasado para predecir su comportamiento futuro probable, entre otras.

Como se ha mencionado, el cálculo de la función de distancia d es muy costoso en la mayoría de los casos y por ello se la utiliza como medida de complejidad en ese modelo. Entonces, para tratar de reducir estos costos han surgido varias técnicas para resolver el problema de consultas por similitud en un número sublineal de cálculos de distancia, generalmente basadas en el *preprocesamiento* de los datos.

Entre las primitivas básicas de las búsquedas por similitud, se encuentra la recuperación de los *k-vecinos más cercanos* a un elemento dado (k -NN(u)). Esta puede definirse como: sea X un conjunto de elementos y d la función de distancia defi-

nida entre ellos, los k -NN(u) son los k elementos en $X - \{u\}$ que tengan la menor distancia a u de acuerdo con la función d . Una variante menos estudiada de este problema, es la búsqueda de los k -vecinos más cercanos de *todos* los elementos de X , *All-k-NN*. Es decir, si $|X| = n$, obtener los *All-k-NN* es calcular los k -NN(u_i) para cada u_i en X ; por supuesto realizando menos de n^2 cálculos de distancia. Algunas soluciones a este problema fueron propuestas y desarrolladas para espacios métricos generales [14, 13], basadas en la construcción del *Grafo de los k-vecinos más cercanos* (k NNG). Éste es un grafo dirigido ponderado que conecta cada elemento del espacio métrico mediante un conjunto de arcos cuyos pesos se calculan de acuerdo a la métrica del espacio en cuestión. El k NNG indexa un espacio métrico, requiriendo una cantidad moderada de memoria, y luego se utiliza en la resolución de las consultas por similitud. El desempeño en las búsquedas por similitud de dicha propuesta supera al obtenido utilizando las técnicas clásicas basadas en pivotes.

Por otro lado, teniendo en la mira la reducción de la cantidad de cálculos de distancias posibles durante una búsqueda, se ha planteado el estudio de un enfoque *aproximado* eficiente para resolver estas consultas por similitud. Este enfoque consiste en permitir una relajación en la precisión de la respuesta a fin de obtener una mejora en la complejidad de la de consulta [16, 3, 17]. El objetivo de la *búsqueda por similitud aproximada* es reducir significativamente los tiempos de búsqueda al permitir algunos errores en el resultado de la consulta. Para ello se provee, además de la consulta, un parámetro de precisión ε que controla cuán lejos queremos el resultado de esta consulta del resultado correcto. Un comportamiento razonable para este tipo de algoritmos es acercarse asintóticamente a la respuesta correcta como ε se acerca a cero. Por lo tanto, el éxito de una técnica de aproximación se basa en la resolución del compromiso calidad/tiempo [4]. Esta alternativa a la búsqueda por similitud “exacta” abarca algoritmos aproximados y probabilísticos.

Lenguajes de Consulta

Si se piensa en una base de datos simplemente como una estructura finita, se pueden utilizar las lógicas para expresar consultas sobre éstas. El empleo de lógicas para expresar consultas (o problemas) da origen a la complejidad descriptiva, que no clasifica a los problemas por la utilización de recursos como el tiempo y el espacio, sino que lo hace según el uso de recursos lógicos tales como el número de va-

riables, cuantificadores, operadores, etc. Existe una relación estrecha entre estos dos tipos de complejidades para clases que se identifican con la computación factible, pero se requiere que el dominio de las estructuras sea ordenado. En ese caso la clase de complejidad P es capturada por FO (*First-Order Logic*) extendida con un operador de punto fijo. Aún así, estas lógicas todavía resultan incompletas, ya que ninguna caracterización lógica de computación factible es conocida para estructuras cuyo dominio no está ordenado. En [6] A. Dawar demuestra que ninguna extensión de la lógica de punto fijo, con un número finito de cuantificadores generalizados, captura la clase de complejidad P; así es importante utilizar diferentes lógicas para separar problemas que, en complejidad clásica son vistos como similares.

A. Dawar [5] también demuestra que ciertos problemas NP completos sobre inequivalencia de autómatas finitos son expresables en el fragmento existencial de la lógica SO^ω mientras que el problema NP completo 3-coloreabilidad no lo es. Nosotros definimos la lógica SO^F [8], una restricción semántica de la SO que exige que las variables de segundo orden se interpreten con relaciones cerradas bajo FO tipos. Demostramos que SO^ω está incluida estrictamente en SO^F , y también que en el fragmento existencial de SO^F podemos expresar con simpleza la propiedad de rigidez que no es posible expresar en SO^ω , es decir, rigidez es la propiedad que separa SO^F de SO^ω . Además, así como el fragmento existencial $\Sigma_1^{1,\omega}$ de la lógica SO^ω captura la clase de complejidad NP_r , contenida en la clase NP, nosotros, con el fragmento existencial $\Sigma_1^{1,F}$ de SO^F , capturamos la clase de complejidad NP^F que posee un problema perteneciente a co-NP. Con lo que conjeturamos que la clase de complejidad NP^F no está incluida en la clase de complejidad NP.

Además definimos una variante del juego de Ehrenfeucht-Fraïssé para la lógica SO^F monádico [7] y demostramos que dicho juego captura la semántica de SO^F monádico. En base a esto se mostró que 2-coloreabilidad no es expresable en SO^F monádico, lo que es curioso ya que 2-coloreabilidad pertenece a la clase de complejidad P y sin embargo no es posible expresarlo en una lógica capaz de discernir si dos elementos en una estructura satisfacen las mismas propiedades de FO. Esto implica que hay un automorfismo que conmuta dichos elementos. Como no se sabe que el problema de establecer automorfismos en una estructura esté en P (la solución por fuerza bruta es exponen-

cial), deberíamos concluir que conocer este tipo de propiedades no basta para resolver 2-coloreabilidad eficientemente.

Por otro lado, también mostramos que un problema de SO^F monádico como “una estructura posee un único FO tipo para elementos” no es expresable en SO monádico existencial, donde si se puede expresar 3-coloreabilidad que se sabe que NP-completo. Es decir, en una lógica donde se pueden expresar problemas que se presumen que no tienen una solución eficiente, no es posible expresar que dados dos elementos cualesquiera hay un automorfismo que los conmuta.

En otra línea de investigación que continúa con la línea estudiada por Dawar en SO^ω , la cual plantea una restricción semántica a la SO, donde la valuación de las variables relacionales para los cuantificadores de segundo orden son cerrados bajo \equiv_k , definimos una nueva lógica de tercer orden (TO), la cual hemos llamado TO^ω . Ésta intenta caracterizar y estudiar clases de complejidad relacionales (temporales) de lógicas de orden superior. Una relación se dice cerrada bajo \equiv_k si todas las tuplas del dominio sobre el que trabaja, que tienen el mismo tipo (satisfacen las mismas formulas de FO^k), están en la relación. Se definió una variación de una máquina relacional no determinística, que denotamos como 3-NRM, donde permitimos relaciones de tercer orden en el *relational store*; ésta nos permitió asociar TO^ω a una clase de complejidad temporal. Esa clase de complejidad fue llamada $NEXPTIME_{3,r}$, como la clase de máquinas 3-NRM que trabajan en tiempo exponencial de acuerdo al tamaño de la entrada. La clase $NEXPTIME_{3,r}$ es exactamente caracterizada por el fragmento existencial de TO^ω [1].

Resultados y Objetivos

Estos estudios, sobre espacios métricos y sobre algunas estructuras de datos particulares, permitirán no sólo mejorar el desempeño de las mismas sino también aplicar, eventualmente, muchos de los resultados que se obtengan a otros MAMs. Nos planteamos considerar los distintos aspectos relacionados al diseño de estructuras de datos que, conscientes de la jerarquía de memorias y de las características particulares de los datos a ser indexados, logren ser eficientes en espacio y en tiempo. Por ello, se busca que los índices se adapten mejor al nivel de la jerarquía de memorias donde se almacenarán. Se espera que nuestras propuestas brinden herramientas de administración eficiente al modelo de bases

de datos métricas y así permitan que el desarrollo de dicho modelo se acerque al que tienen los modelos tradicionales de base de datos.

Respecto de los lenguajes de consulta se continuará analizando la expresividad de distintas extensiones y posibles restricciones de SO y TO, para lograr caracterizar la clase de las consultas computables sobre bases de datos no convencionales.

Actividades de Formación

Dentro de esta línea de investigación se forman alumnos y docentes-investigadores en:

Doctorado en Cs. de la Computación: un investigador finalizó su tesis sobre bases de datos métricas. Otro integrante está realizando su tesis sobre la expresividad de la lógica como lenguaje de consulta.

Maestría en Cs. de la Computación: un investigador de la línea desarrolla su tesis sobre búsqueda por similitud aproximada, a finalizar este año.

Maestría en Informática: un alumno de la Universidad Nacional de San Juan está desarrollando su tesis sobre un índice dinámico para búsquedas por similitud aproximadas en memoria secundaria.

Trabajo Final de Ingeniería Civil en Computación: un alumno de la Universidad de Talca finalizó su trabajo de fin de carrera sobre el diseño de un nuevo índice dinámico para memoria secundaria, basado en combinar pivotes con *Lista de Clusters*.

Referencias

- [1] J. Arroyuelo and J. Turull Torres. The Existential Fragment of Third Order Logic and Third Order Relational Machines. XX Congreso Argentino de Cs. de la Computación: 324–333. 2014.
- [2] E. Chávez, K. Figueroa, and G. Navarro. Effective proximity retrieval by ordering permutations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9):1647–1658, 2008.
- [3] E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, 2001.
- [4] P. Ciaccia and M. Patella. Approximate and probabilistic methods. *SIGSPATIAL Special*, 2(2):16–19, 2010.
- [5] A. Dawar. A Restricted Second Order Logic for Finite Structures. *Information and Computation*, 143, 154–174, 1998.
- [6] A. Dawar. Feasible Computation through Model Theory. Ph.D. thesis, University of Pennsylvania, 1993.
- [7] A. L. Grosso. Tesis Doctoral, SO^F : Una lógica donde las variables de relación se interpretan con uniones de FO tipos, Universidad Nacional de San Luis.
- [8] A. L. Grosso and J. M. Turull Torres. A Second-Order Logic in which Variables Range over Relations with Complete First-Order Types. XXIX International Conference of the Chilean Computer Science Society. IEEE 270 – 279, 2010.
- [9] G. Navarro. Searching in metric spaces by spatial approximation. *The Very Large Databases Journal*, 11(1):28–46, 2002.
- [10] G. Navarro and N. Reyes. Dynamic spatial approximation trees. *Journal of Experimental Algorithmics*, 12:1–68, 2008.
- [11] G. Navarro and N. Reyes. Dynamic spatial approximation trees for massive data. *Procs. of 2nd International Conference on Similarity Search and Applications*, 81–88. IEEE, 2009.
- [12] G. Navarro and N. Reyes. Dynamic list of clusters in secondary memory. *Proc. of 7th International Conference on Similarity Search and Applications*, LNCS 8821, 94–105. 2014.
- [13] R. Paredes. *Graphs for Metric Space Searching*. PhD thesis, University of Chile, Chile, 2008.
- [14] R. Paredes, E. Chávez, K. Figueroa, and G. Navarro. Practical construction of k -nearest neighbor graphs in metric spaces. In *Proc. 5th Workshop on Efficient and Experimental Algorithms*, LNCS 4007, 85–97. 2006.
- [15] R. Paredes and N. Reyes. Solving similarity joins and range queries in metric spaces with the list of twin clusters. *Journal of Discrete Algorithms*, 7(1):18–35, 2009.
- [16] H. Samet. *Foundations of Multidimensional and Metric Data Structures (The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling)*. Morgan Kaufmann Publishers Inc., 2006.
- [17] P. Zezula, G. Amato, V. Dohnal, and M. Batko. *Similarity Search: The Metric Space Approach (Advances in Database Systems)*. Springer-Verlag, 2005.

Geometría Computacional y Bases de Datos

Maria Gisela Dorzán, Susana Esquivel, Edilma Olinda Gagliardi,

Pablo Palmero y Maria Teresa Taranilla

Departamento de Informática
Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis, Argentina
{mgdorzan, esquivel, oli, prpalmero, tarani}@unsl.edu.ar

Gregorio Hernández Peñalver

Departamento de Matemática Aplicada
Facultad de Informática
Universidad Politécnica de Madrid, España
gregorio@fi.upm.es

Resumen

La línea de investigación “Geometría Computacional y Bases de Datos” enmarcada en el proyecto “Tecnologías Avanzadas de Bases de Datos” vincula el estudio de las disciplinas Bases de Datos, Geometría Computacional y Metaheurísticas, utilizando los métodos y las herramientas provistas para la resolución de problemas orientados a optimización.

Palabras clave: Bases de Datos, Geometría Computacional, Metaheurísticas, Bases de Datos Espaciales y Espacio-Temporales.

Contexto

El proyecto “Tecnologías Avanzadas de Bases de Datos” de la Universidad Nacional de San Luis tiene como objetivo el estudio de bases de datos avanzadas, y en este marco se investiga el diseño y desarrollo de herramientas que permiten administrar de manera eficiente sistemas de bases de datos no estructuradas.

En el proyecto coexisten tres líneas de investigación, orientadas al desarrollo de nuevos modelos para administrar y recuperar información almacenada en repositorios de datos no estructurados, donde los escenarios de exploración requieren modelos tales como las bases de datos espaciales, bases de datos de

texto, bases de datos de imágenes, bases de datos de sonidos, espacios métricos, entre otros.

En este sentido, en el proyecto “Tecnologías Avanzadas de Bases de Datos” se desarrollan actividades vinculadas al tratamiento de objetos de diversos tipos, estructurados y no estructurados, útiles en diversos campos de aplicación, por ejemplo, robótica, computación gráfica, visión artificial, sistemas de información geográfica, diseño asistido por computadora, computación móvil, motores de búsqueda en internet, entre otras, y que se relacionan en tales bases de datos.

Así, surge el estudio de modelos como las bases de datos espaciales y bases de datos espacio-temporales y la construcción y manipulación de diferentes objetos y estructuras geométricas de utilidad en diversas áreas de aplicación. En particular, las estructuras geométricas que se estudian deben cumplir con propiedades deseables y algunos de los problemas relacionados con la optimización de las mismas son problemas NP-duros, por tanto en la búsqueda de soluciones aproximadas se recurre a técnicas metaheurísticas.

En la línea de investigación “Geometría Computacional y Bases de Datos” se vinculan temáticas que surgen de las disciplinas Bases

de Datos, Geometría Computacional y Metaheurísticas.

El trabajo de investigación se desarrolla en forma conjunta con investigadores afines de proyectos locales y de universidades extranjeras (Universidad Politécnica de Madrid - España, Universidad Veracruzana - México, entre otras) mediante convenios de cooperación institucional. En este trabajo, se presentan temas de interés junto con las propuestas más recientes de trabajo.

Introducción

La Geometría Computacional estudia el diseño de algoritmos para resolver problemas geométricos. En esta disciplina se identifican conceptos, propiedades y técnicas que apuntan al desarrollo de algoritmos eficientes. Esto conduce al análisis y estudio de estructuras de datos geométricas y problemas relacionados con dichas estructuras [BCKO08]. Los problemas se estudian desde un punto de vista combinatorio o algorítmico. Los aspectos algorítmicos de los problemas aparecen en la búsqueda de soluciones, exactas o aproximadas, para una estructura geométrica particular. En este contexto aparecen problemas que pertenecen a la clase de complejidad NP, o bien, son problemas para los cuales no se conocen algoritmos eficientes que permitan alcanzar soluciones exactas. En determinadas circunstancias interesa obtener soluciones a tales problemas, aunque sean aproximadas.

Algunos de los problemas refieren a ciertas configuraciones geométricas obtenidas a partir de un conjunto de puntos u objetos en el plano y se busca optimizar propiedades que miden la calidad de dichas configuraciones. Los objetivos están orientados básicamente a construir estructuras geométricas que satisfagan ciertas propiedades vinculadas a problemas de optimización para los cuales hay estudios acerca de su naturaleza NP-dura. Para ello, se propone la aplicación de técnicas metaheurísticas. Una metaheurística es una metodología de alto nivel que puede ser utilizada como una estrategia que guía las heurísticas subyacentes para resolver

problemas de optimización específicos, combinando inteligentemente diferentes conceptos de diversos campos. Estos métodos son simples de implementar y han demostrado poder encontrar, de forma eficiente, buenas soluciones para problemas de optimización NP-duros [MF04] [BFM97].

Los problemas en los que se trabajaron corresponden a estructuras geométricas, obtenidas a partir de un conjunto de puntos u objetos en el plano, para las cuales se busca optimizar alguna propiedad que permita valorar la calidad de dichas estructuras. Los criterios de optimización consisten en minimizar o maximizar alguna propiedad de las estructuras geométricas.

Entre los problemas de optimización estudiados destacamos: la Triangulación de Peso Mínimo (*Minimum Weight Triangulation, MWT*) y la Pseudo-Triangulación de Peso Mínimo (*Minimum Weight Pseudo-Triangulation, MWPT*). Dichos problemas minimizan la suma de las longitudes de las aristas proveyendo así una medida de calidad de la estructura. La complejidad del cálculo de MWT fue uno de los problemas abiertos más estudiados en Geometría Computacional hasta que se demostró que la construcción de MWT es un problema NP-duro [MR06]. Con respecto a la complejidad de MWPT aún no está resuelta. Existe un 12-aproximación de una pseudo-triangulación que puede ser calculada con complejidad de $O(n^3)$ y una aproximación de $O(\log n w(\text{MST}))$ de una pseudo-triangulación de peso mínimo en un tiempo de $O(n \log n)$ donde $w(\text{MST})$ es el peso del Árbol de Expansión Mínimo (*Minimum Spanning Tree, MST*) el cual es un subconjunto de la estructura obtenida. [GL07].

Otro problema en estudio es la Triangulación de Dilación Mínima (*Minimum Dilation Triangulation, MDT*) donde la dilación mide la calidad de conexión entre puntos de la triangulación. La dilación entre un par de puntos, u y v , de una triangulación T se define como la razón entre el camino más corto entre u y v en T y la distancia Euclídea entre dicho par de puntos. La máxima dilación entre todos los pares de puntos en T se llama la dilación de

T ($\Delta(T)$). La mejor posible dilación de cualquier triangulación de un conjunto de puntos S se conoce como dilación de S ($\Delta(S)$). La complejidad del problema de encontrar $\Delta(S)$ para un conjunto de puntos S se desconoce y por lo tanto, el trabajo se centra en el desarrollo de algoritmos aproximados para encontrar triangulaciones de buena calidad de dilación mínima. Los problemas de optimización relacionados a triangulaciones y pseudo-triangulaciones son de interés debido a su utilidad en diversos campos: detección de colisiones, visibilidad y vigilancia, problemas de rigidez de estructuras, entre otros.

La complejidad de los problemas antes mencionados justifica la utilización de técnicas metaheurísticas que permitan obtener soluciones aproximadas a las óptimas.

Línea de investigación, Desarrollo e Innovación

En la línea de investigación se propone el estudio de optimizaciones de estructuras geométricas relacionadas con las bases de datos mencionadas, el desarrollo de herramientas para la visualización de estructuras geométricas y aplicaciones vinculadas a las bases de datos espacio-temporales. En particular, los problemas de interés en la línea de investigación son aquellos vinculados al diseño de índices espacio-temporales para resolver en forma integral consultas espacio-temporales y su vinculación con la problemática de objetos móviles; y al tratamiento de problemas geométricos NP-duros. Para estos últimos, en la búsqueda de soluciones se utilizan las técnicas metaheurísticas que proporcionan estrategias adecuadas para la obtención de soluciones aproximadas de múltiples problemas de optimización.

Como objetivos específicos de estudio podemos enumerar los siguientes:

- Estudio de configuraciones geométricas generales de puntos en el plano, que cumplan medidas de calidad como: peso, dilación, área, entre otras. Optimización de triangulaciones y pseudo-triangulaciones, considerando medidas

de calidad mínimas o máximas, aplicando técnicas metaheurísticas y otras estrategias algorítmicas.

- Diseño y desarrollo de herramientas para la generación, visualización y manipulación de diferentes configuraciones geométricas de conjuntos de puntos en el plano.

- Desarrollo de aplicaciones utilizando herramientas de Geometría Computacional y Bases de Datos Espacio-Temporales.

Resultados y Objetivos

Para la resolución de problemas de optimización de triangulaciones y pseudo-triangulaciones de peso mínimo se aplicaron las técnicas metaheurísticas: Optimización basada en Colonia de Hormigas (*Ant Colony Optimization, ACO*) y Recocido Simulado (*Simulated Annealing, SA*), técnicas determinísticas Voraces (*Greedy*) y Triangulación de Delaunay. Se llevó a cabo el estudio, adecuación y evaluación experimental de las técnicas metaheurísticas mencionadas para la búsqueda de triangulaciones y pseudo-triangulaciones que cumplan la propiedad. Se diseñaron generadores de instancias de problema para ser utilizados en la evaluación experimental. Dichas colecciones se encuentran disponibles en el sitio de la línea de investigación (www.dirinfo.unsl.edu.ar/bd2/GeometriaComp/).

Los resultados obtenidos para los problemas MWT y MWPT utilizando la técnica metaheurística ACO fueron publicados en [DGLH11b] [GDLH11] [DGLH12]. Se llevó a cabo una evaluación experimental y análisis de los resultados obtenidos con la técnica Recocido Simulado [DGLH11a] [DGLH11c].

Para el problema MDT todavía no se conoce un algoritmo que lo resuelva en tiempo polinomial y tampoco se ha demostrado que sea NP-duro. Las técnicas aplicadas para este problema fueron: Greedy, Local Search, Iterated Local Search, Simulated Annealing y Random Local Search. Para cada estrategia se propuso un conjunto de operadores adecuados. Debido a la complejidad que implica la puesta a punto de los parámetros de técnicas metaheurísticas se utilizó Optimización de

Parámetros Secuencial (*Sequential Parameter Optimization, SPO*) para el ajuste de los parámetros requeridos por Simulated Annealing. Se realizó un análisis experimental en el cual se compararon dichos algoritmos con otras técnicas, como por ejemplo, Delaunay. Se crearon las instancias de prueba, ya que para estos problemas no se encontraron disponibles ningún tipo benchmark con el cual comparar nuestros resultados. Las conclusiones fueron afirmadas desarrollando un estudio estadístico aplicando diferentes test estadísticos y métodos de visualización [DLMH14].

Los resultados obtenidos en el tratamiento de estos problemas se plasmaron en dos tesis de doctorado.

Por otra parte, se desarrolló una herramienta para la generación y visualización de triangulaciones, pseudo-triangulaciones y poligonizaciones de conjuntos de puntos en el plano. Además, se finalizó con la implementación de una aplicación en el ámbito de la Salud para el seguimiento de focos epidémicos utilizando base de datos espacios-temporales y herramientas de Geometría Computacional. El desarrollo de las herramientas mencionadas se plasmó en dos trabajos finales de Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas [PDG13], [GPDGT14].

Como trabajo futuro, considerando los problemas sobre configuraciones geométricas en estudio, se pretende continuar con el estudio de los problemas de optimización considerando otras técnicas metaheurísticas, adecuadas para su resolución.

Formación de Recursos Humanos

La formación del grupo de trabajo en la Universidad Nacional de San Luis, se consolida con actividades de cooperación mutua e intercambio recíproco de información científica, tecnología y desarrollo de nuevos conocimientos con investigadores locales y de otras universidades.

Entre las actividades más destacadas se pueden mencionar:

- Formación de recursos humanos plasmada dos Tesis Doctorales (UNSL), tres Licenciados en Ciencias de la Computación (UNSL).
- Obtención de becas de investigación tipo I y tipo II de CONICET.
- Realización y dirección de pasantías de investigación con docentes de otras universidades.
- Actividades de formación académica, dictado de cursos de posgrado y de especialización.
- Actividades de divulgación científica, conferencias, publicaciones en congresos y revistas en el ámbito nacional e internacional.

Por lo expuesto anteriormente, la línea de investigación “Geometría Computacional y Bases de Datos” tiene como propósito continuar con el trabajo de investigación para consolidar y alimentar líneas de estudio en la disciplina que permitan fortalecer y afianzar la formación de recursos humanos locales, proponiendo actividades de formación académica, de investigación, desarrollo, y otras actividades académico-científicas vinculantes para que ello represente un aporte en la comunidad científica.

Referencias

- [BFM97] Bäck T., Fogel D., Michalewicz Z. *Handbook of Evolutionary Computation*. IOP Publishing Ltd and Oxford University Press. 1997.
- [BCKO08] de Berg, M., Cheong O., van Kreveld, M., Overmars, M., *Computational Geometry: Algorithms and Applications*. 3rd edition, Springer-Verlag, Heidelberg, 2008.
- [DGLH11a] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Approaches for MWT and MWPT Problems*. XVII Congreso Argentino de Ciencias de la Computación 2011 (CACIC 2011), 2011
- [DGLH11b] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Using ACO metaheuristic for MWT problem*. XXX International Conference of the Chilean Computer Science Society. ISBN 978-0-7695-4689-6. Chile. 2011

- [DGLH11c] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Metaheuristic approaches for MWT and MWPT Problems*. XIV Encuentros de Geometría Computacional. Páginas: 79-82. 2011
- [DGLH12] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Approximations on Minimum Weight Triangulations and Minimum Weight Pseudo-Triangulations using Ant Colony Optimization Metaheuristic*. Fundamenta Informaticae. ISSN: 0169-2968 (Print), 1875-8681 (Online). Volume 119, number 1, pp 1-27.
- [DLMH14] M. G. Dorzán, M. G. Leguizamón, Efrén Mezura-Montes, G. Hernández Peñalver *Approximated algorithms for the Minimum Dilution Triangulation Problem*. Journal of Heuristics. DOI 10.1007/s10732-014-9237-2. Print ISSN 1381-1231. Online ISSN 1572-9397. Publisher Springer US. 2014.
- [GDLH11] Gagliardi E. O., Dorzán M. G., Leguizamón M. G. y Hernández Peñalver. G.; *Approximations on Minimum Weight Pseudo-Triangulation problem using Ant Colony Optimization*. XXX International Conference of the Chilean Computer Science Society. ISBN 978-0-7695-4689-6. Chile. 2011
- [GL07] Gudmundsson J., Levcopoulos C.; *Minimum weight pseudo-triangulations*. Computational Geometry. Theory and applications. Elsevier Vol. 38- pages 139-153, 2007.
- [GPDGT14] Guasch, M.M; Piergallini, M.R ; Dorzán, M.G.; Gagliardi, E.O.; Taranilla, M.T.; *“Una herramienta para el análisis y seguimiento de focos epidémicos”* en Anales del 17° Concurso de Trabajos Estudiantiles en 43 Jornadas Argentinas de Informática. Páginas: 35-45- Universidad de Palermo, Buenos Aires, Argentina, 2014.
- [MR06] Mulzer W., Rote G. *Minimum weight triangulation is NP-hard*. In Proceedings of the 22nd Annual ACM Symposium on Computational Geometry. 2006.
- [MF04] Michalewicz Z., Fogel D., *How to Solve It: Modern Heuristics*, Springer, 2004.
- [PDG13] Palmero, P.R., Dorzán, M. G., Gagliardi E.O., *Una Herramienta para la Manipulación de Configuraciones Geométricas*, 42° Jornadas Argentinas de
- Informática e Investigación Operativa - 16° Concurso de trabajos de fin de carrera. Universidad Nacional de Córdoba, Córdoba, Argentina. 2013. ISSN: 1850-2946.

Plataforma Web para acceder a diferentes Sistemas de Gestión de Bases de Datos Relacionales

Muñoz Roberto Miguel, Maldonado Calixto, Damiano Luis Esteban, Romero
María Soledad, Cuevas Juan Carlos, Quinteros Sergio Ramón, Guevara
Andrea, Carrasco Agustín

Grupo de Investigación, Desarrollo y Transferencia de Sistemas de Información
Departamento Ingeniería en Sistemas de Información
Facultad Regional Córdoba / Universidad Tecnológica Nacional
Maestro Marcelo López esq. Cruz Roja Argentina – Ciudad Universitaria - Córdoba
0351 - 4686385

{robertmunioz, calixtomaldonado, luis.damiano, romeroma.soledad, juancarloscue, ser.quinteros,
andrezza77, asermax}@gmail.com

Resumen

El propósito del proyecto es estudiar, investigar, diseñar y proponer una Plataforma Web para colaborar en la capacitación y ejercitación del lenguaje SQL, accediendo a bases de datos relacionales.

Es una herramienta de ayuda a la práctica personal, donde el usuario encuentra un entorno con estructuras de datos, tipificación de sentencias y consignas a resolver, para recibir una devolución que le permite crecer en aprendizaje.

Las especificaciones del software de interfaz y del middleware se tomarán a partir de necesidades de usuarios, que inicialmente serán los estudiantes de la cátedra de Gestión de Datos, del tercer nivel de la Carrera de Ingeniería en Sistemas de Información.

El middleware ofrecerá a los usuarios la conexión a distintos motores de base de datos, permitiendo aplicar el lenguaje SQL según los distintos productos que puedan incluirse.

Además brindará un módulo de gestión para usuarios administradores, con privilegio de edición, así pueden

definir estructuras de trabajo, ejecutar scripts con tablas que contengan datos y agregar consignas para ejecutar sentencias SQL.

La arquitectura será web, para que el usuario no necesite instalar software cliente en su equipo y le baste con un browser.

Palabras clave: Base de datos, modelo relacional, SQL, Plataforma web, Software como Servicio, SaaS

Contexto

El proyecto ha sido homologado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional, con un tiempo de realización desde el 1 de enero de 2016 y por 24 meses. La Unidad Científico-Tecnológica donde se desarrolla es el GIDTSI (Grupo de Investigación, Desarrollo y Transferencia de Sistemas de Información), de la UTN, Facultad Regional Córdoba.

El origen del proyecto se centra en observaciones y encuestas realizadas en la

Cátedra de Gestión de Datos, perteneciente al tercer nivel de la Carrera de Ingeniería en Sistemas de Información, donde un alto porcentaje del cronograma de cursado se destina al estudio y ejercitación del Lenguaje SQL.

El grupo de trabajo está conformado inicialmente por siete docentes investigadores de la carrera, con la colaboración de un graduado y dos estudiantes becarios, de la misma carrera.

Los docentes se desempeñan en cátedras relacionadas a las siguientes temáticas: Programación, Base de Datos y Sistemas de Información.

El equipo de investigación tiene experiencia profesional y en docencia, pero se destaca que mantienen continuidad en proyectos de investigación relacionados a la temática:

- TecnoDB - Administrador de Base de Datos Relacional (2007) [1]
- PROMETEO - Desarrollo de un método y una herramienta para el aprovechamiento de Metadatos de Base de Datos Relacionales (2010) [2]
- Análisis y aplicación de metodologías para la generación de consultas complejas utilizando esquemas OLAP (2010) [3]
- Generador Automático de Modelos de Datos Normalizados en Bases de Datos Relacionales (2015) [4]

Cabe destacar que los integrantes: Ing. Roberto Muñoz, Ing. Calixto Maldonado y el Lic. Luis Damiano, participaron junto a los Ingenieros Enrique Reinoso y Maximiliano Abrutsky, en la escritura del libro "Base de Datos" [5].

Introducción

La temática abordada está relacionada con las bases de datos relacionales, que si bien inicialmente sólo se las utilizaba para persistir datos desde las aplicaciones de las empresas [6], hoy son utilizadas para la toma de decisiones a partir de la

exploración de las bases de datos, con la Minería de Datos [7, 8].

Se considera un tema relevante, dado su impacto en la cantidad de horas hombre destinadas a la resolución de temas inherentes al manejo de datos, en el entorno de bases de datos relacionales. Más aún si se considera que es una tarea componente de la mayor parte de los proyectos de software y que de ella dependen otras como el desarrollo y el testing. Todo desarrollador de aplicaciones debe conocer SQL, para embeber el código necesario en sus programas y así acceder/manipular los datos que solicita un usuario final [9, 10, 11].

Los docentes de la Cátedra de Gestión de Datos observaban que no todos los estudiantes lograban ejercitar en sus equipos personales y sólo lo hacían en el momento de la ejercitación práctica guiada por el auxiliar docente.

Buscando lograr que el entrenamiento desde la cátedra sea eficaz, los docentes fueron proponiendo e incorporando medidas a través del tiempo:

- Actualización permanente de la Guía de Trabajos Prácticos de Gestión de Datos, con ejercicios combinados y revisión permanente.
- Creación de script con la generación del modelo de datos "Empresa" y la carga de datos, para poder ejercitar en Oracle. Se instaló en las aulas del Laboratorio de Ingeniería en Sistemas de Información (LabSis), para que los estudiantes resolvieran ejercicios.
- Con el tiempo se decidió ampliar el alcance del script del modelo, para ser ejecutado en los motores: MySQL, PostgreSQL, DB2 y SQL Server. Estos fueron publicados en el repositorio del LabSis para los estudiantes.
- Posteriormente se creó el script con la estructura y los datos que figuran en la bibliografía de la asignatura, para ser

usado en SQLServer. Ese script también fue publicado.

- Un docente generó una máquina virtual con: sistema operativo, DBMS y los scripts de generación de estructuras y carga de datos.

Esto muestra que el tema de la ejercitación de SQL se analiza de manera permanente, pretendiendo motivar a los estudiantes para una mejor ejercitación y dejarles buen conocimiento de SQL.

A pesar de lo señalado, se continúan detectando inconvenientes en los procesos de aprendizaje de los estudiantes.

La consecuencia de no practicar lo suficiente es que no logran la detección temprana de errores u omisiones, que sólo se alcanza con la resolución de ejercicios.

En el año 2013 la cátedra de Gestión de Datos implementó una encuesta, para conocer las opiniones de los estudiantes y lo relacionado a aspectos de la práctica de SQL. Fue respondida por 219 estudiantes, de los 350 inscriptos. Sólo el 8% de los encuestados indicaron que la práctica era Excelente. Se entiende que los estudiantes también reconocen que la práctica podría mejorarse. Además se verificaron inconvenientes que suelen tener:

- Incompatibilidades entre sistema operativo y DBMS. Es común que los estudiantes consigan el producto a instalar y el proceso de instalación suele no ser exitoso.

- Inconvenientes entre versiones de software de base de datos y script, por ejemplo problemas en formato de fechas.

- Falta de experiencia en la instalación de software. Un estudiante respondió: “Instalé SQL Server pero no pude ingresar por el usuario”.

- Dificultad para conseguir el motor adecuado para el sistema operativo.

Estas situaciones también se producen cuando los estudiantes preparan un examen final, debido al tiempo que suele

pasar entre el fin de cursado del estudiante y el examen.

A partir de esto se generó la idea de crear una plataforma web, accesible a toda hora y desde cualquier sitio con internet, que permita realizar prácticas de SQL y que devuelva un feedback, sobre si es correcta la solución o en caso contrario brinde mensajes claros que indiquen la oportunidad de mejora, colaborando así en el aprendizaje del lenguaje.

Existen plataformas que ofrecen el acceso web, las cuales se describen a continuación y se señalan diferencias:

- Navicat: Software comercial que permite administrar bases de datos de manera remota. El objetivo es netamente comercial y principalmente orientado a Administradores de Bases de Datos. Registrado por PremiumSoft. [12]

- RAT (Relational Algebra Translator): Software que traduce Álgebra Relacional a SQL. El RAT fue desarrollado en la Universidad Nacional de Costa Rica. RAT necesita un software cliente y depende del Sistema Operativo [13].

También se realizaron consultas sobre software que permita mejorar el aprendizaje del lenguaje, pero no fueron encontrados.

Líneas de Investigación, Desarrollo e Innovación

En el GIDTSI una de las líneas de investigación es la Ingeniería en Sistemas de Información y de Software, incluyendo proyectos en calidad, métricas, mantenimiento, trazabilidad, requerimientos, bases de datos, seguridad, relacionados con las áreas de conocimiento: Sistemas de Información, Gestión Ingenieril y Programación. En ese contexto el proyecto desarrolla temáticas relacionadas a las bases de datos, con un importante desarrollo de programación.

En una de las líneas de investigación de las posibles soluciones a la problemática, se pueden realizar aprendizaje a través de ontologías de consultas SQL. Considerando como una ontología [14, 15] al conjunto finito de posibles construcciones SQL que satisfacen el mismo resultado esperado a un ejercicio propuesto. Se podría lograr una ontología para cada ejercicio de práctica, compuesta por los posibles comandos SQL que satisfacen la solución, que el sistema utilizará para validar la solución del usuario. El hecho de poseer los comandos SQL que satisfacen la solución al problema planteado permite guiar al usuario con el conjunto de datos que se deberían obtener en el ejercicio, esta posibilidad de acceder a una vista del resultado esperado (datos), es un vehículo que permite visualizar, a través de ingeniería inversa, cuál debiera ser el comando SQL para llegar al resultado. Tratando las posibles soluciones a un problema de SQL como ontologías, tanto para los comandos, como para el conjunto de datos resultado, el camino correcto sería a través del análisis de patrones, complementado con minería de texto.

El diseño y construcción del prototipo iniciará orientado a la estructura de datos para almacenar el metadato y a una interfaz sencilla para los estudiantes, incluyendo la ejecución de sentencias del sublenguaje DML. Resuelto ello, deberá orientarse al entorno de administración para los docentes.

Resultados y Objetivos

El resultado esperado es que la plataforma permita al usuario:

- Acceder a un intérprete SQL, para ejecutar sentencias SQL, accediendo a una estructura de datos ya poblada.
- Ejecutar sentencias tipificadas, para que los estudiantes elijan.

- Observar el resultado de la ejecución, pudiendo listar y copiar las filas resultantes, cuando es una consulta.
- Recibir mensajes de la plataforma cuando haya un error en la sintaxis.
- Validar el resultado a través de mensajes sobre la cantidad de filas y columnas devueltas por una consulta.
- Recibir mensajes sobre si el resultado es el esperado o qué error se detecta contra el resultado esperado.

El Objetivo General planteado es: Estudiar, investigar, diseñar y proponer una Plataforma Web para colaborar en la capacitación y ejercitación del lenguaje SQL, con bases de datos relacionales. Dicho objetivo se logrará cuando el equipo de investigación logre:

- a) Determinar requerimientos a partir de las necesidades del profesor y del estudiante.
- b) Evaluar las alternativas elegibles, ventajas y desventajas de modelos de persistencia de datos en la web.
- c) Identificar alternativas de integración de diferentes motores de bases de datos.
- d) Apoyar el proceso de selección de lenguajes de programación y tecnología de persistencia de los datos utilizando herramientas multicriterio de evaluación.
- e) Diseñar y desarrollar un prototipo middleware experimental en entorno web.
- f) Gestionar la publicación del middleware con herramientas de integración continua.
- g) Analizar los logs de uso para encontrar relaciones ocultas en el uso de la herramienta. Ejemplo: éxitos y fracasos.
- h) Buscar patrones de comportamiento de los usuarios que permitan definir indicadores de acierto y errores en la escritura.
- i) Construir las ontologías que contribuyan a la sistematización de la observación de los resultados.
- j) Realizar pruebas de validación del esquema propuesto con profesionales.
- k) Medir resultados con el uso de la aplicación en distintos empresas de software de la ciudad de Córdoba.

Formación de Recursos Humanos

El equipo de trabajo está constituido por: un director de proyecto, un codirector, cinco docentes investigadores, dos estudiantes becarios y un graduado.

Seis de los docentes comparten su trabajo en la cátedra de Gestión de Datos, donde se trata la temática de Bases de Datos y específicamente el Lenguaje SQL. El director del proyecto es el director de dicha cátedra, quien desarrollará su tesis de la Maestría en Ingeniería en Sistemas de Información en el seno del proyecto.

El estudiante Agustín Carrasco desarrollará la Práctica Supervisada [16] centrada en la construcción del prototipo y el estudiante Mauricio Spalletti ya ha tenido experiencia como becario en UTN.

Referencias

- [1] Gastañaga, I., Maldonado, C., Martínez Spessot, C. I., & Hintermeister, E. (2006). TecnoDB una Base de Datos Relacional y Prometeo un método de aprovechamiento de Metadatos y Generador de Consultas. In VIII WICC.
- [2] Marciszack, M., Maldonado, C., Martínez Spessot, C. I., Muñoz, R., Navarro, A., Peretti, J. P., & Roggero, L. (2009). Prometeo: una herramienta para el aprovechamiento de metadatos de base de datos relacionales. In XI WICC.
- [3] Vaca, P. A., & Abrutsky, M. (2009). Herramienta para consultas complejas orientada a usuarios finales. In XI WICC.
- [4] Paz Menvielle, M. A., Cuevas, J. C., Damiano, L. E., Muñoz, R., & Quinteros, S. (2013, June). Generador automático de modelos de datos normalizados en bases de datos relacionales. In XV WICC.
- [5] Reinoso, E.; Maldonado, C.; Muñoz, R.; Damiano, L.; Abrutsky, M.- Bases de Datos-Edit.AlfaOmega Editores-Argentina - 2012- ISBN: 978-987-1609-31-4.
- [6] Date, C.J.Introducción a los Sistemas de Bases de Datos. Volumen 1. Séptima Edición. México, 2001. Edit. Addison Wesley Longman, Inc. ISBN: 0-201-38590-2.
- [7] Bigus, J. P. Data Mining with Neural Networks: Solving Business Problems from Application Development to Decision Support. 1996. McGraw-Hill. ISBN:0-07-005779-6
- [8] Riquelme, J., Ruiz, R y Gilbert, K. Minería de Datos: Conceptos y Tendencias. Artículo publicado por Revista Iberoamericana de Inteligencia Artificial. ISSN: 1137-3601. <http://www.redalyc.org/articulo.oa?id=92502902>. Último acceso: 11/12/2015, a las 14:35 hs.
- [9] Introducción a los Sistemas de Bases de Datos. Volumen 1. Séptima Edición. México, 2001. Edit. Addison Wesley Longman, Inc. ISBN: 0-201-38590-2
- [10] Elmasri, R. y Navathe, S. Fundamentos de Sistemas de Bases de Datos. 5ta Edición, 2007 Edit. Pearson. ISBN: 978-84-7829-085-7.
- [11] Silberschatz y otros- Fundamentos de Bases de Datos - Quinta Edición- Edit. Mc Graw Hill- Estados Unidos - 2006 - ISBN: 84-481-4644-1.
- [12] <http://www.navicat.com/es/>. Último acceso: 25/02/2016, a las 12:30 hs.
- [13] <http://www.slinfo.una.ac.cr> Último acceso: 20/12/2015, a las 17:45 hs.
- [14] Mario Bunge – Tratado de Filosofía – Ontología I – Editorial Gedisa – Barcelona 2011 – ISBN: 978-84-9784-196-2
- [15] María Jesús Lamarca Lapuente– Ontologías.<http://www.hipertexto.info/documentos/ontologias.htm> . Último acceso: 12/12/2015, a las 17:00 hs.
- [16] UTN -FRC - Prácticas Supervisadas <http://www.institucional.frc.utn.edu.ar/sistemas/Areas/Alumnos/PracticasSupervisadas.asp>. Último acceso: 20/02/2015, a las 10:15 hs.

Big Data y su impacto sobre las comunidades

Fernando Emmanuel Frati¹, Jose Texier^{1,2}, Fernanda Carmona¹, Daniel Robins¹, Alberto Riba¹, Cristian Rios¹

¹ Departamento de Ciencias Básicas y Tecnológicas, Universidad Nacional de Chilecito
9 de Julio 22, Chilecito, La Rioja, Argentina
{fegrati, jtexier, fbcarmona, drobins, ariba}@undec.edu.ar,
riosbourne555@gmail.com

² Universidad Nacional Experimental del Táchira
Venezuela

Resumen

La necesidad de administrar grandes volúmenes de información no estructurada (heterogénea) es un tema que en la actualidad está en crecimiento. El auge se debe a la enorme producción de información digital, tanto la información producida en Internet como la cuantiosa información producida por las instituciones (llámense, empresas u organismos estatales). La cuestión es que en su mayoría tal cantidad de información no es administrada correctamente lo cual conlleva pérdidas o complicaciones en el funcionamiento. El propósito de la creación de esta línea de investigación es analizar el contexto de oportunidades presentes en la región de Chilecito y ofrecer alternativas a las diferentes problemáticas presentes desde la perspectiva del Big Data. Para tal fin, esta línea de I/D/I corresponde al diseño, desarrollo e implementación de proyectos que fortalecerán la investigación y las diferentes labores relacionadas con las necesidades de la región sobre la base del recurso humano presente en la Universidad Nacional de Chilecito (UNdeC).

Palabras clave: *Big Data, Chilecito, procesamiento de información*

Contexto

Esta línea de I/D/I corresponde al diseño, desarrollo e implementación de proyectos que fortalecerán la investigación y las diferentes labores relacionadas con las necesidades de la región sobre la base del recurso humano presente en la Universidad Nacional de Chilecito (UNdeC).

En la actualidad, se están desarrollando ideas-proyectos a través de una tesina de grado sobre el tema de Big Data aplicados a la Vitivinicultura. Para la próxima convocatoria (2016) del programa “Financiamiento para el Estímulo y Desarrollo de la Investigación Científica y Tecnológica” de la Secretaría de Ciencia y Tecnología en la UNdeC (FiCyT - UNdeC) se presentará un proyecto sobre Big Data en los procesos de elaboración de vinos. También se está desarrollando un proyecto de análisis social mediante tweets en determinados temas, por ejemplo, elecciones presidenciales de la Argentina del 2015. Es importan-

te destacar que el estudiante involucrado en el proyecto ha sido beneficiado con una beca de Estímulo a las Vocaciones Científicas, convocatoria EVC 2015 (CIN).

Por otra parte, la UNdeC cuenta con la estructura tecnológica y de RRHH necesarios para ejecutar los proyectos que surjan de la línea de investigación.

Introducción

La Universidad Nacional de Chilecito se ubica en uno de los centros urbanos más importante de la región de los valles áridos del noroeste argentino. Históricamente la investigación y desarrollo de la UNdeC ha sido impulsada por las necesidades del sector agrícola. Sin embargo, las políticas de la universidad en términos de inversión y vinculación tecnológica llevadas a cabo en los últimos años ha derivado en la instalación de dos importantes laboratorios para la región que ofrecen un enorme potencial de trabajo.

Este trabajo explora las oportunidades de I+D+I en Big Data[1], entendiendo que estas actividades deben ser realizadas en su contexto, beneficiándose del trabajo interdisciplinario como un importante motor de desarrollo regional. Se denomina Big Data a aquellos problemas que requieren procesar grandes volúmenes de datos en poco tiempo [2, 3, 4]. Por ello, el procesamiento y posterior análisis debe ser realizado (preferiblemente) en tiempo real para poder mejorar la toma de decisiones con base en la información generada. Las características del Big Data se centran en las tres V:

- Volumen de grandes volúmenes de datos.
- Variedad de diversos tipos de fuentes de datos, ya sean estructurados o no estructurados.

- Velocidad de los datos, es decir, la frecuencia de las actualizaciones de estas grandes bases de datos.

A partir del contexto descrito, esta propuesta analiza las potencialidades de la zona en conjunto con el RRHH presente en la UNdeC. Por ello, a continuación se proponen algunas ideas dirigidas a aplicar estos conceptos en la región:

Actividades agroindustriales: el clima de Chilecito se caracteriza por la extrema aridez, con grandes amplitudes térmicas, lluvias anuales medias de 200 milímetros, concentradas en época estival; fuerte insolación anual, frecuentes vientos desecantes y baja humedad atmosférica. Pese al marcado déficit hídrico típico de la región de los valles áridos, lleva adelante una intensa actividad agrícola industrial. La mayor concentración de cultivos en el subsector de fruticultura lo tienen el olivo con 11.000 hectáreas, la Vid con 6.500 hectáreas y el Nogal con 2.800 hectáreas, los cuales se comercializan a nivel local, regional, nacional e internacional. El proceso de industrialización de algunos cultivos como el de la vid se lleva adelante en 15 bodegas que se distribuyen en la ciudad y distritos del Departamento Chilecito. La fabricación de aceite de oliva está en pleno crecimiento, dando cuenta de un reducido número de productores que utilizan tecnología de avanzada para su elaboración.

Laboratorios de Altura y de Alta Complejidad: el laboratorio de altura es el primero de América de esta clase, y constituye un proyecto de excelencia en las actividades científicas-tecnológicas. Se encuentra a 5200 metros sobre el nivel del mar, en las Sierras del Famatina, de fácil acceso, excelentes condiciones atmosféricas y con la Universidad al pie de las sierras, lo que aporta el valor agregado de disponibilidad de recursos humanos formados

(y en formación). Este laboratorio permite hacer mediciones que no se pueden hacer al nivel del mar en campos como medicina, biología, astronomía, física, etc. Por otro lado el LAC presta los siguientes servicios a la comunidad: análisis de suelos, análisis de aguas para riego, análisis de aguas para consumo, análisis microbiológico de agua, análisis de efluentes, análisis en vías de implementación. Cuenta con una gran cantidad de instrumental de laboratorio y de campo, y actualmente está en proceso la adquisición de un secuenciador de ADN, un Secuenciador Genómico y un Microscopio Electrónico de Barrido de alta resolución, lo que permitirá ampliar los servicios ofrecidos. Ambos laboratorios cuentan con una importante cantidad de profesionales e investigadores de otras disciplinas e instrumental muy especializado. En consecuencia, existe una necesidad de trabajo con grandes volúmenes de datos y procesos complejos con requerimientos de tiempo real.

Oportunidades: se han desarrollado varias tesis de grado (dirigidas por integrantes de este trabajo) afines a esta línea, involucrando tecnologías de microcontroladores y redes de sensores para resolver distintos problemas relacionados a la captura de información, automatización de procesos de control y predicción de variables asociadas a los procesos agrícolas. Sin embargo, estos proyectos se han desarrollado como sistemas a medida destinados solamente a la organización beneficiaria de cada proyecto. Por otro lado se están realizando reuniones de trabajo con las personas a cargo de ambos laboratorios en búsqueda de espacios de colaboración. Se propone llevar a gran escala estos proyectos, coordinando la participación de múltiples empresas de la región. Es de esperar que esto nos permita evaluar técnicas existentes e implementar desarrollos experimentales para cla-

sificar, ordenar, jerarquizar y analizar información sobre grandes volúmenes de datos heterogéneos. Como objetivo principal nos proponemos implementar un repositorio de datos públicos a disposición de los investigadores de la región, aportando a cuestiones específicas sobre agrodatos de vid, olivo y nogales para Chilecito y zonas de influencia. Estos “Data Set” públicos nos habilitarán a pensar nuevos temas derivados de esta línea de trabajo.

La propuesta del proyecto permitirá establecer las capacidades necesarias con las que debería contar una base de datos de información masiva, tanto desde la perspectiva de almacenamiento y técnicas de indexación, como de distribución de las consultas, escalabilidad y rendimiento en ambientes heterogéneos, unido con el análisis de las oportunidades de aplicación del Big Data a la región.

Líneas de Investigación, Desarrollo e Innovación

- Bases de datos, Minería de datos
- Cloud Computing[11], Arquitecturas paralelas
- Análisis social web[12]
- Simulación
- Internet de las cosas
- Agromática, Vitivinicultura, Genética
- Repositorios institucionales y bibliotecas digitales, Análisis semántico de la información

Para llevar a cabo el proyecto propuesto, se plantean las siguientes actividades:

- Exploración de oportunidades de aplicación del Big Data en la región y relacionado con el recurso humano existente en la UNdeC.
- Obtención de datos. Por ejemplo, datos correspondientes a compras realizadas por personas, datos médicos de pacientes, estructuras de ADN, datos de redes sociales[5, 6].
- Modelado de los datos de forma relacional, documental, gráfica, clave-valor, y familia-columnas.
- Traslado de cada modelo a los diferentes motores de base de datos SQL y NoSQL [7, 8, 9].
- Análisis de los datos capturados, transformados y almacenados[10].
- Visualización de los datos.
- Pruebas mediante consultas complejas en cada uno de los motores para cada uno de los modelos desarrollados.
- Determinar si existen casos donde una estructura no influye en la eficiencia de las consultas bajo un determinado paradigma y si los resultados son comprensibles.
- Diseñar e implementar Data Set según las necesidades de la UNdeC.

Resultados y Objetivos

Objetivos

- Generar un entorno de desarrollo sobre tecnología Big Data.
- Consolidar un grupo de investigación multidisciplinario en la UNdeC.
- Fomentar, incentivar y difundir las tareas de investigación.

- Mejorar la formación de recursos humanos con capacidades de investigación y desarrollo.
- Generar y establecer Data Sets públicos con información regional para ofrecerlos a los investigadores de la universidad.

Resultados esperados

- Formación del recurso humano.
- Identificación de las diferentes oportunidades de proyectos para la región.
- Diseño de un conjunto de estrategias para abordar las oportunidades en la región.
- Generación de un marco de trabajo que permite a la comunidad contar con una estructura para solucionar sus problemas, enfocados sobre el dominio de datos masivos.

Formación de Recursos Humanos

El equipo de trabajo está formado por docentes de las carreras Ingeniería en Sistemas y Licenciatura en Sistemas de la UNdeC (acreditadas por CONEAU), dos doctores especializados en repositorios institucionales, bibliotecas digitales, desarrollo de software, cómputo paralelo y tecnología grid. Otra docente está definiendo su tesis de Maestría en Informática. También participa un alumno avanzado de grado. En otras palabras, se cuenta con un recurso humano con habilidades y formación académica en las diversas áreas de la propuesta, asegurando la concreción de la línea. Adicionalmente, se destaca que dos están categorizados en el programa de incentivos.

Los integrantes son docentes de las asignaturas Programación I, Sistemas I, Arquitecturas Paralelas, Teoría de la Computación, Bases de Datos I y II, y Herramientas de Ingeniería de Software. Estas asignaturas contemplan la aprobación mediante la participación en proyectos de investigación, por lo que pueden surgir nuevos trabajos en esta línea.

Referencias

- [1] D. López, “Análisis de las posibilidades de uso de Big Data en las organizaciones,” *Universidad de Cantabria, Santander, España*, 2012.
- [2] A. Rajaraman, J. D. Ullman, J. D. Ullman, and J. D. Ullman, *Mining of massive datasets*. Cambridge University Press Cambridge, 2012, vol. 1.
- [3] B. Dana and C. Kate, “‘Critical Questions for Big Data’,” *Information, Communication and Society*, vol. 15, p. 5, 2012.
- [4] C. Lynch, “Big data: How do your data grow?” *Nature*, vol. 455, no. 7209, pp. 28–29, Sep. 2008.
- [5] N. Antonopoulos and L. Gillam, *Cloud Computing: Principles, Systems and Applications*. Springer Science & Business Media, Jul. 2010.
- [6] W. Hall, D. D. Roure, and N. Shadbolt, “The evolution of the Web and implications for eResearch,” *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1890, pp. 991–1001, Mar. 2009. [Online]. Available: <http://rsta.royalsocietypublishing.org/content/367/1890/991>
- [7] E. Serrano and C. A. Iglesias, “Validating viral marketing strategies in Twitter via agent-based social simulation,” *Expert Systems with Applications*, vol. 50, pp. 140–150, May 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417415008234>
- [8] B. Li, B. F. Liu, and N. X. Liu, “The Big Data how to Influence the Development of E-Government,” in *Advanced Materials Research*, vol. 989. Trans Tech Publ, 2014, pp. 4717–4722. [Online]. Available: <http://www.scientific.net/AMR.989-994.4717>
- [9] H. Ramírez and J. F. Herrera, “Un viaje a través de bases de datos espaciales,” *NoSQL: Redes de ingeniería, Univ. Distrital Francisco J de Caldas, Bogotá, Colombia*, vol. 4, pp. 35–47, 2013.
- [10] H. G. del Busto and O. Y. Enríquez, “Bases de datos NoSQL,” *Revista Telemática*, vol. 11, no. 3, pp. 21–33, 2013.
- [11] C. Nance, T. Losser, R. Iype, and G. Harmon, “Nosql vs rdbms-why there is room for both,” in *Proceedings of the Southern Association for Information Systems Conference*, 2013, pp. 111–116.
- [12] E. E. Schadt, M. D. Linderman, J. Sorenson, L. Lee, and G. P. Nolan, “Computational solutions to large-scale data management and analysis,” *Nature Reviews Genetics*, vol. 11, no. 9, pp. 647–657, Sep. 2010.

Sentiment Analysis para la clasificación de noticias financieras en los Mercados Argentinos. Un modelo híbrido de POST enriquecido semánticamente.

Juan Pablo Braña, Alejandra M.J. Litterio y Alejandro Fernández
Centro de Altos Estudios en Tecnología Informática- Facultad de Tecnología Informática -
U.A.I
Av. Montes de Oca 745 - (C1270AAH) Ciudad Autónoma de Buenos Aires,
República Argentina
{juan.brana, alejandra.litterio, alejandroa.fernandez}@uai.edu.ar

Resumen

El proyecto de investigación en curso, que aquí se presenta, propone un modelo híbrido enriquecido semánticamente, en el cual aplicar un *etiquetador morfosintáctico* con el fin de identificar cómo una determinada secuencia de palabras, a partir de una estructura sintáctica, refleja un indicador de sentimiento, esto es, clasificar una cláusula en positivo, negativo o neutro, dentro de un contexto específico, en nuestro caso particular los Mercados Financieros Argentinos. Con el propósito de llevar a cabo este estudio recolectamos, analizamos y clasificamos opiniones extraídas de usuarios de Twitter, comentarios de blogs especializados en finanzas, artículos periodísticos en economía y finanzas – que constituirá nuestro corpora ampliado–, aplicando principios y técnicas de Sentiment Analysis y Machine Learning.

Palabras clave: Sentiment Analysis, Machine Learning, Mercados Financieros Argentinos, Computational Linguistics, POS Tagging.

Contexto

El presente proyecto, cuyo inicio es Marzo 2016, se desarrolla en el Centro de Altos Estudios en Tecnología Informática (CAETI) dependiente de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana (UAI). Se enmarca dentro de la línea de Algoritmos y Software y continúa las investigaciones iniciadas en el Proyecto “Modelo de Sentiment Analysis para la clasificación de noticias en tiempo real”. Es financiado y evaluado por la Secretaría de Investigación de la Universidad. Cuenta con la participación de docentes y alumnos de la Maestría en Tecnología Informática y de la Diplomatura en Análisis de Datos para Negocios, Finanzas e Investigación de Mercados.

Introducción

Como hemos observado en nuestro estudio anterior, identificar y extraer información subjetiva de comentarios en las redes sociales (principalmente, Twitter) para el estudio de los mercados financieros constituye, en la actualidad, un tema de radical interés para desarrollar herramientas analíticas que brindan la posibilidad a analistas

financieros de usar estas opiniones para mejorar la precisión de las predicciones del mercado. De la misma manera hemos hecho referencia al surgimiento de nuevas formas de realizar transacciones financieras mediante el *Trading Algorítmico* (AT) o trading automático y el estrecho vínculo con *Sentiment Analysis*, siendo lo más notable de esta tecnología la capacidad de un algoritmo de detectar e interpretar semánticamente una noticia o acontecimiento en el mismo instante que se está produciendo y así conocer su correlación con los movimientos de precios con el Mercado de Valores. Ahora bien, todo esto ha sido explicado considerando aspectos esenciales de la clasificación de textos determinando su tendencia positiva o negativa desde su *estructura o composición semántica contextualmente dependiente*. Así, hemos mostrado cómo estos aspectos semánticos del lenguaje en un contexto o dominio específico son fundamentales para generar modelos de *Machine Learning* con buenos índices de certeza para clasificar tweets relacionados a finanzas en los mercados argentinos, en positivos o negativos.

Por otra parte, creemos que es imprescindible dar continuidad al estudio y realizar el tratamiento de nuestra corpora incorporando el nivel sintáctico, si bien las tecnologías o modelos basados en semántica han revolucionado no sólo la forma en que integra y se le asigna una valoración a la información sino también a los resultados en el análisis de la polaridad.

El modelo de procesamiento de lenguaje natural (PLN) que aquí proponemos consiste en aplicar un *etiquetador morfosintáctico* con el fin de identificar cómo una determinada secuencia de palabras, a partir de una estructura sintáctica, refleja un indicador de sentimiento, esto es, clasificar una

cláusula en positivo, negativo o neutro— nos extendemos a textos completos que contienen una carga valorativa como artículos periodísticos en finanzas y mercados (“noticias duras”), comentarios en blogs especializados en finanzas y notas de opinión general asociadas a la temática que estamos analizando. En una primera etapa, evaluaremos los etiquetadores ya existentes y su aplicabilidad a nuestra corpora. En una segunda fase determinaremos si es necesario realizar alguna mejora incorporando estructuras sintagmáticas y paradigmáticas de mayor complejidad a aquellas ya disponibles. Hacia la etapa final de la investigación, propondremos un *modelo híbrido enriquecido semánticamente*, es decir, que nos permita no sólo etiquetar sino también establecer una correspondencia con el *lexicón FSAL* (Financial Spanish Dictionary of Affect) y verificar si existe un avance en cuanto al desempeño (performance) de los algoritmos de clasificación mediante la aplicación de este nuevo modelo.

En la literatura existente, encontramos diversos trabajos que documentan el uso de etiquetadores morfosintácticos (POS Tagger, en inglés) para incrementar el desempeño en la clasificación de sentimiento u opiniones de comentarios de consumidores sobre un producto o reseña literaria (Patel y Chang, 2014; Jagtap and Pawar, 2013, Nicholls and Song, 2009). En estudios comparados de etiquetadores MBT, QTAG y MXPSOT de corpora en español, así como en mejora y adaptación de métodos de etiquetado (Martí et al., 2006). Cabe mencionar, también, *El Grial*, una interfaz computacional que permite la realización de anotaciones morfosintácticas en textos planos en lengua española además de la consulta en forma de base de datos de los corpora allí reunidos (Parodi, 2006; Venegas, 2008).

En el ámbito de mercados y finanzas, y más específicamente en el dominio de las noticias en financieras, se encuentran los trabajos de Devitt y Ahmad (2007) para explicar la relación entre el contenido afectivo de los textos y su impacto en el mercado.

Líneas de Investigación, Desarrollo e Innovación

En la actualidad los Mercados Financieros Argentinos se encuentran en una etapa de plena innovación en cuanto a Trading Algorítmico. Cabe señalar que tanto el ROFEX, el Merval como el MAV ya cuentan con las plataformas necesarias para operar con esta nueva tecnología. De modo que, todo un ecosistema que une inversores y analistas financieros con desarrolladores, científicos, investigadores en el ámbito académico están trabajando de manera colaborativa con el propósito de avanzar en este nuevo escenario.

El presente trabajo se enmarca en la investigación de aplicación de técnicas y algoritmos de minería de datos en bases de datos y pretende dar un enfoque y soporte académico a toda la comunidad sea esta financiera como académica. En este estudio, en particular sentamos las bases para aplicar las herramientas de Sentiment Analysis, teniendo en cuenta que las noticias financieras son uno de los pilares fundamentales a la hora de la toma de decisiones por parte de los inversores.

Resultados y Objetivos

Pregunta Problema: Un modelo combinado de análisis morfosintáctico enriquecido semánticamente y enmarcado en un contexto económico-financiero en los mercados argentinos podría arrojar mejores resultados al momento de determinar la polaridad de un texto, que si

consideráramos estos mismos modelos de manera autónoma e independiente del contexto de aplicación.

Objetivo General:

Avanzar en los aspectos sintácticos de Sentiment Analysis, como el etiquetado morfosintáctico (en inglés, Part Of Speech Tagging), brindándole mayor complejidad lingüística a nuestros modelos y una aproximación más formal a la Inteligencia Artificial aplicada a las Finanzas, especialmente en los Mercados Argentinos.

Objetivos Específicos:

- a. Adentrarse en los aspectos Sintácticos de la Lingüística Computacional creando modelos de Sentiment Analysis basados en POST (Part of Speech Tagging).
- b. Entrenar diversos Algoritmos Estadísticos para la clasificación de Noticias en “Positivas”, “Negativas” o “Neutras”, de acuerdo a los modelos del punto anterior.
- c. Crear señales de compra/venta de instrumentos financieros basados en los resultados obtenidos en nuestros modelos y correlacionar dichos resultados con valores en tiempo real de los mercados argentinos.
- d. Desarrollar diferentes Indicadores de Sentimiento y correlacionarlos con diferentes indicadores financieros, como por ejemplo Volatilidad o Volumen.
- e. Estudiar las posibilidades de crear un modelo híbrido entre los resultados obtenidos en la etapa anterior del proyecto y los nuevos objetivos, más específicamente un modelo híbrido entre los aspectos Semánticos y Sintácticos del Lenguaje en modelos de Sentiment Analysis.
- f. Ampliar el corpus en referencia de la etapa anterior, esto es, aquel que se utilizó en el desarrollo de la investigación 2013-2015. De modo que, no sólo se trabajará

con tweets sino también con artículos periodísticos financieros, comentarios de blogs especializados en finanzas y notas de opinión general relacionadas a la temática propuesta en este proyecto.

Metodología de Trabajo

El enfoque teórico-metodológico del presente trabajo se basa en aplicar herramientas de Natural Language Processing y Machine Learning desarrolladas en los lenguajes de programación R y Python y descrito en los siguientes puntos:

1. Con un corpus constante, es decir, mantenemos el corpus del proyecto anterior en una fase inicial, esto es 800 tweets clasificados en POSITIVOS y NEGATIVOS de manera manual por una comunidad de expertos en finanzas.
2. Se realiza un estudio de diferentes herramientas Open Source de análisis del lenguaje: Freeling, openNLP, NLTK y Stanford Log Linear Part-Of-Speech Tagger para evaluar los resultados que se obtienen en conjunto y con cada uno de ellos.
3. Se crean modelos de Sentiment Analysis basado en Part of Speech Tagging (Etiquetador Morfosintáctico) y los clasificadores Naive Bayes y Random Forest. Los mismos se implementan en la plataforma estadística R y la librería CARET.
4. Se comparan los resultados entre la fase semántica obtenida anteriormente y la fase morfosintáctica de nuestra presente investigación.
5. Se evalúa adecuar las herramientas Open Source utilizadas en base a los resultados obtenidos. Esto es, mejorar manualmente las herramientas analizadas en el punto (2).
6. Se propone la creación de un modelo combinado que incluya los niveles de

semántica-morfosintáctica para enriquecer nuestro modelo de Sentiment Analysis aplicado a finanzas.

Resultados Esperados

En una primera instancia esperamos que las clasificaciones arrojadas por nuestros algoritmos, sin un tratamiento previo de las herramientas descritas en el punto (2), al que se refiere en la sección Metodología de Trabajo, tengan una performance inferior a los resultados obtenidos a nivel semántico en el estudio llevado a cabo durante 2013-2015. Sin embargo, consideramos que estos resultados pueden tener un desempeño superior si perfeccionamos manualmente los sistemas analizadores mencionados.

Por último, en un estadio más avanzado, nos proponemos crear un modelo combinado que muestre un mayor grado de precisión al integrar los niveles de semántica y morfosintaxis, que incrementen el grado de fiabilidad de las anotaciones morfosintácticas además de producir etiquetas en español que sean transparentes y acertadas en su nominación, y que superen el resultado de los modelos aplicados por separado.

Formación de Recursos Humanos

El equipo del proyecto, multidisciplinario, se compone principalmente de docentes de la Diplomatura en Análisis de Datos para Negocios, Finanzas e Investigación de Mercado, y la Maestría en Tecnología Informática así como expertos del área de Lingüística y Finanzas. Por su parte, señalamos que el proyecto contará con la participación de alumnos avanzados de la Maestría en Tecnología Informática, quienes llevan a cabo su pasantía de investigación al tiempo que identifican temas en los que puedan desarrollar su

tesis. Además se cuenta con la colaboración de alumnos de la mencionada Diplomatura.

Referencias

Braña, J.P.; Litterio, A.; Camós, C. y Fernández, A. (2015). Modelo de Sentiment Analysis para la clasificación de noticias en tiempo real en el Mercado de Valores de Buenos Aires. En *XVII Workshop de Investigadores en Ciencias de la Computación*, Salta. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/45547>

Devitt, A. and Ahmad, K. (2007). "Sentiment Polarity Identification in Financial News: A Cohesion-based Approach", *Annual Meeting of the Association of Computational Linguistics*, Prague, Czech Republic.

Enriquez, F.; Troyano, J.; Cruz, F. y Ortega, J. (2006). Ampliación automática de corpus mediante la colaboración de varios etiquetadores. En *Procesamiento del Lenguaje Natural*, Nro. 37, pp. 6 -11.

Genereux, M., Poibeau, T. and Koppel, M. (2008). "Sentiment analysis using automatically labeled financial news items". *LREC 2008 Workshop on Sentiment Analysis: Emotion, Metaphor, Ontology and Terminology*, Marrakech, Morocco.

Hernández, S., Miranda, S., Jiménez, E., Villaseñor, E., Tellez, S. y Graff, M. (2015). "Minería de opinión en blogs financieros para la predicción de tendencias en mercados bursátiles". *Research in Computing Science*, Vol.92, pp. 101-109.

Jagtap, V. S. and Pawar, K. (2013). "Analysis of different approaches to Sentence-Level Sentiment Classification", *International Journal of Scientific Engineering and Technology*, Vol. 2, pp. 164-170.

Martí, J. et al. (2006). Adaptación del Método de Etiquetado No Supervisado TBL. En *Procesamiento del Lenguaje Natural*, Nro. 37, pp. 3-5.

Morales de Jesús, V. M. (2014). *Utilización de expresiones de actitud para el Análisis de Sentimientos*. Tesis de Licenciatura, Benemérita Universidad Autónoma de Puebla, México.

Nicholls, C. and Song, F. (2009). "Improving sentiment analysis with Part-of-Speech weighting". Department of Computation & Information Science, University of Guelph, Guelph, ON, Canada.

Parodi, G. (2006). "El Grial: Interfaz Computacional Para Anotación e Interrogación de Corpus en Español". En *Revista de Lingüística Teórica y Aplicada*, 44 (2), II Sem., pp. 91-115.

Patel, N. D. et al. (2014). "Selecting Best Features Using Combined Approach in POS Tagging for Sentiment Analysis", *International Journal of Computer Science and Mobile Computing*, Vol.3 Issue.3, pp. 425-430.

Peifeng, L., Qiaoming, Z. and Wei, Z. (2011). "A Dependency Tree Based Approach for Sentence-Level Sentiment Classification". *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 12th ACIS International Conference*. pp, 166-171.

Venegas, R. (2008). Interfaz Computacional de Apoyo al Análisis Textual: "El Manchador de Textos". En *Revista de Lingüística Teórica y Aplicada*, 46 (2), II Sem., pp. 53-79.

Comportamiento de Bases de Datos No Relacionales en Entornos Distribuidos

Ana Lía Carabio¹, Marcelo G. Benedetto¹, Marcelo A. Falappa²

¹Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos
Monseñor Tavella 1424 – Concordia, Entre Ríos (3200) - Tel.: +54(0345)4231406
[anacar,marben}@fcad.uner.edu.ar](mailto:{anacar,marben}@fcad.uner.edu.ar)

²Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur
Avenida Alem 1253 - Bahía Blanca (B8000CPB) - Tel.: +54(0291)4595135
mfalappa@cs.uns.edu.ar

Resumen

La estructura de un sistema de información típico actual consta, en general, de programas de aplicación distribuidos, un Sistema de Gestión de Base de Datos (SGBD) y una red que permite entregar la información desde y hacia los distintos usuarios.

Dentro de los SGBD actuales han surgido las bases de datos NOSQL (Not Only SQL) y los Sistemas de Almacenamiento de Datos Masivos (*Big Data Storage Systems*), que almacenan la información en forma distribuida y permiten una mayor escalabilidad.

El tiempo utilizado en la comunicación entre los distintos nodos afecta la performance de la base de datos y, por consiguiente, el tiempo de respuesta al usuario, poniendo en evidencia la importancia del comportamiento de la red en el rendimiento de un sistema de información.

En este trabajo de investigación se buscará analizar el comportamiento de una base de datos no puramente

relacional, y el comportamiento de la red de comunicaciones, con el fin de evaluar el rendimiento a medida que se actualiza la información en un entorno distribuido.

Palabras clave: Bases de Datos, Sistemas Distribuidos, Redes de Comunicación, Medidas de Performance.

Contexto

Este trabajo se desarrolla dentro del Proyecto de Investigación y Desarrollo PID 7042 “Estudio Comparativo y Análisis de Rendimiento de los Lenguajes de Manipulación de Datos en Bases de Datos Orientadas a Objetos y Bases de Datos Objeto-Relacionales”[1], cuyo período de ejecución será desde noviembre de 2014 a noviembre de 2017, en el marco de un Acuerdo de Colaboración Académico-Científico entre la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos (UNER) y el Instituto de Ciencias e Ingeniería de la Computación (ICIC) del Departamento de Ciencias e Ingeniería de la Computación

(DCIC) de la Universidad Nacional del Sur (UNS).

Uno de los objetivos del proyecto apunta a establecer comparaciones en el rendimiento de sistemas desarrollados en lenguajes orientados a objetos que interactúan con diversos modelos de bases de datos.

Además, este proyecto prioriza la formación de recursos humanos para investigación en la Facultad de Ciencias de la Administración de la UNER, especializados en la línea de investigación denominada “Ingeniería de Software y Lenguajes de Programación” establecida por Res. 25/11 del Consejo Directivo.

Introducción

Las redes de computadoras y las bases de datos son utilizadas masivamente por las aplicaciones de hoy en día, dado que la disponibilidad y acceso a la información se ha convertido en una herramienta indispensable para la toma de decisiones.

La estructura de un sistema de información típico actual consta, en general, de programas de aplicación distribuidos, un SGBD y una red que permite entregar la información desde y hacia los distintos usuarios.

Dentro de los SGBD, el modelo relacional ha sido el más utilizado, a pesar de ciertas limitaciones. En la actualidad, existen extensiones de estos sistemas que incorporan los conceptos de tipos complejos y orientación a objetos conformando los SGBDOR *Sistemas de Gestión de Base de Datos Objeto-Relacionales* (SGBDOR) y los *Sistemas de Gestión de Base de Datos Orientados*

a Objetos (SGBDOO) [2]. También han surgido las bases de datos NOSQL (*Not Only SQL*) [2], las cuales permiten mayor escalabilidad que los sistemas tradicionales [3], y los Sistemas de Almacenamiento de Datos Masivos (*Big Data Storage Systems*). Estos sistemas han emergido en compañías y organizaciones que almacenan grandes cantidades de información en forma distribuida, entre las cuales podemos citar Google, Yahoo, Amazon, Facebook, etc.

En cuanto a la tecnología de redes de computadoras, la misma promueve una forma de trabajo que procura, en general, evitar la centralización; mientras que uno de los objetivos del uso de las bases de datos es la necesidad de integrar los datos y proveer mecanismos que controlen el correcto acceso a los mismos [4].

En los sistemas distribuidos los datos se encuentran almacenados en varios servidores, y los clientes pueden acceder a ellos por medio de una red de comunicación [5, 6]. Cuando el SGBD y/o los datos se encuentran alojados en sitios remotos, las consultas a la base de datos pueden generar que la información atraviese varias redes y/o dispositivos (nodos) hasta llegar a destino, lo que incide directamente en el rendimiento.

Independientemente de la red utilizada, la transferencia de información a través de cualquier canal de comunicación requiere de un tiempo que dependerá del ancho de banda del canal, de la longitud y sobrecarga del enlace, de la velocidad y eficiencia de la red, del número de nodos, entre otros [6, 7]. El tiempo utilizado en la comunicación constituye un retardo que afecta la performance de la base de

datos y, por consiguiente, el tiempo de respuesta al usuario [8].

Las bases de datos relacionales y objeto-relacionales fueron concebidas, inicialmente, para instalaciones centralizadas, aunque conceptualmente pueden implementarse de manera distribuida [8]. Cuando las aplicaciones se ejecutan en un entorno distribuido, tanto el programador de aplicación como el usuario del sistema deberían independizarse de aspectos físicos tales como: cantidad de nodos de la red, topología de la misma, réplicas de los datos, y/o fragmentación de las relaciones. No obstante, si nos abstraemos totalmente de la red y no consideramos que los recursos están físicamente ubicados en varias computadoras [4], se corre el riesgo de caer en los “famosos” supuestos de la computación distribuida, que terminan siendo falsos y generan grandes problemas a largo plazo: la red es confiable, la latencia es cero, el ancho de banda es infinito, la red es segura, la topología no cambia, existe un solo administrador, el costo de transporte es cero y la red es homogénea [9].

El rendimiento de un SGBD distribuido se ve afectado por el número de sitios en los que dicho sistema está distribuido, así como también por el grado de replicación de los datos, entre otros parámetros [8,10,12]. Por ejemplo, en una base de datos relacional tradicional, las relaciones se pueden fragmentar horizontalmente (por tuplas), verticalmente (a través de la descomposición en subesquemas), y combinado.

En particular, las bases de datos NOSQL permiten una replicación entre nodos más

simple que la de las bases de datos relacionales, utilizando principalmente dos técnicas (replicación y sharding), de las que derivan otros modelos de distribución. La replicación duplica los datos en múltiples nodos, utilizando el modelo maestro-esclavo o el modelo peer-to-peer, mientras que sharding divide (fragmenta) la información en varios nodos. Estas técnicas, con sus ventajas y desventajas, también pueden combinarse entre sí, generando un esquema más complejo [11].

Líneas de Investigación, Desarrollo e Innovación

En la actualidad han cobrado importancia las bases de datos no puramente relacionales, caracterizadas, principalmente, por su almacenamiento distribuido y su fácil escalabilidad. En esta línea, se buscará analizar el comportamiento de bases de datos no relacionales del tipo NOSQL con distintos esquemas de distribución, con la finalidad de evaluar su rendimiento a medida que se actualiza la distribución de la información.

A su vez, la distribución de información calificada, como pueden ser los datos multimediales, se encuentra con ciertos obstáculos que presentan las redes de comunicaciones para soportar este tipo de datos. En particular, la transmisión de ciertos datos requiere un ancho de banda determinado y no deberían perder continuidad. Por ejemplo, cuando se transfiere audio y/o video mediante *streaming*, es necesario que la información tenga continuidad de

reproducción. En este sentido, se intentará formular y/o adaptar protocolos de comunicación que permitan garantizar que la información transferida a través de la red tenga la fluidez esperada.

Resultados y Objetivos

Dada la incidencia que indudablemente tiene el comportamiento de la red sobre el rendimiento global de un sistema de información y de una base de datos distribuida no puramente relacional en particular, se hace necesario evaluar el desempeño de la red y posteriormente determinar la incidencia de dicho comportamiento sobre el rendimiento de la base de datos. Para ello se prevé:

- Seleccionar e instalar distintas bases de datos del tipo NOSQL, con distintos modelos de distribución, sobre una red con al menos dos o tres computadoras personales con procesadores de varios núcleos, y un número incremental de máquinas virtuales.
- Evaluar el desempeño de la red en distintas instancias de trabajo, utilizando como indicadores, inicialmente, la tasa de transferencia (*throughput*), el retardo y la pérdida de paquetes. En caso que la investigación así lo requiera, se estimará la incorporación de otras variables de medición, y/o determinar preferencias entre las variables ya medidas.
- Evaluar el comportamiento de las distintas bases de datos NOSQL instaladas y determinar la incidencia del comportamiento de la red sobre el rendimiento de dichas bases de datos.

Finalmente, se estudiarán los protocolos distribuidos existentes y, si fuera necesario, se reformularán y/o generarán nuevos protocolos con el fin de lograr un mejor comportamiento.

Los resultados de esta investigación dependerán de: las variables de medición contempladas, los datos manipulados por las bases de datos no relacionales, la cantidad de lecturas y escrituras realizadas en la misma, y la cantidad de mensajes de control transferidos a través de la red.

Formación de Recursos Humanos

Como parte del actual proyecto de investigación se espera que uno de los docentes investigadores, y que es autor de este artículo, complete su Tesis de Magister en Redes de Datos en la Facultad de Informática de la Universidad Nacional de La Plata. A su vez, se buscará formar nuevas sublíneas de investigación relacionadas a este proyecto, así como también la formación de nuevos alumnos en los posgrados dictados en el ámbito de la UNER y de la UNS.

Referencias

[1] Benedetto, Marcelo G., Carabio, Ana Lía R., Alvez, Carlos E., Fernández, Miguel, Etchart, Graciela, Cabrera, Sergio A., Benítez, Horacio D., Falappa, Marcelo A, Martínez, Diego C. & Cobo, M. Laura (2015). *Selección de lenguajes orientados a objetos para un estudio comparativo y análisis de rendimiento*. En **XVII Workshop de Investigadores**

en Ciencias de la Computación, WICC'2015.

[2] Elmasri, Ramez, & Navathe, Shamkant B. (2015). *Fundamentals of Database Systems*. 7th. Edition. **Addison Wesley**.

[3] Cattell, Rick (2011). *Scalable SQL and NoSQL data stores*. **ACM SIGMOD Record**, 39(4), 12-27.

[4] Özsu, M. Tamer, & Valduriez, Patrick (2011). *Principles of Distributed Database Systems*. Third Edition. **Springer Science & Business Media**.

[5] Kurose, James F. & Ross, Keith W. (2012). *Computer Networking: A Top-Down Approach*. Sixth Edition. **Pearson**.

[6] Sahu, Amir. K., & Hemrajani, Naveen (2012). *An Analysis of Distributed Computer Network Administration*. **International Journal of Computer Technology and Applications**, 3(2): 660-667.

[7] Gámiz Caro, Juan, & Martínez García, Herminio (2008). *El retardo del mensaje en sistemas de control distribuidos a través de Ethernet estándar*. In la **Quinta Conferencia Internacional de la Facultad de Ingeniería Eléctrica**: 1-7.

[8] Silbertschatz, Abraham. & Korth, Henry. Sixth Edition (2010). *Database System Concepts*. **McGraw-Hill Education**.

[9] Rotem-Gal-Oz, Arnon. (2006). *Fallacies of distributed computing explained*. <http://www.rgoarchitects.com/Files/fallacies.pdf>.

[10] Maabreh, Khaled S. (2011). *An Analyzing Study of the Distributed Database System Parameters*. Technical Report. Faculty of Science and Information technology. Zarqa University. Jordan, Al Zarqa.

[11] Pérez Blanco, Carlos. (2013). *NoSQL databases in cross-platform development*.

[12] Garcia-Molina, Hector, Ullman, Jeffrey D. & Widom, Jennifer (2008). *Database Systems: The Complete Book*. Second Edition. **Pearson**.

Indexación y Búsqueda sobre Datos no Estructurados

Norma Herrera, Darío Ruano, Paola Azar, Susana Esquivel

Departamento de Informática
Universidad Nacional de San Luis, Argentina
{nherrera, dmruano, epazar, esquivel}@unsl.edu.ar

Anabella De Battista

Departamento Ingeniería en Sistemas de Información
FRCU, Universidad Tecnológica Nacional
Entre Ríos, Argentina
debattistaa@frcu.utn.edu.ar

Abstract

Las bases de datos actuales han incluido la capacidad de almacenar datos no estructurados tales como imágenes, sonido, texto, video, etc. La problemática de almacenamiento y búsqueda en estos tipos de base de datos difiere de las bases de datos clásicas, dado que no es posible organizarlos en registros y campos, y aun cuando pudiera hacerse, la búsqueda exacta carece de interés. Es en este contexto donde surgen nuevos modelos de bases de datos capaces de cubrir las necesidades de almacenamiento y búsqueda de estas aplicaciones. Nuestro interés se basa en el diseño de índices eficientes para estas nuevas bases de datos.

1 Contexto

El presente trabajo se desarrolla en el ámbito de la línea Técnicas de Indexación para Datos no Estructurados del Proyecto Tecnologías Avanzadas de Bases de Datos (22/F014), cuyo objetivo es realizar investigación básica en problemas relacionados al manejo y recu-

peración eficiente de información no tradicional.

2 Introducción

La mayoría de los administradores de bases de datos actuales están basados en el modelo relacional, presentado por Edgard F. Codd en 1970. Bajo el modelo relacional, cada elemento de la base de datos puede ser almacenado como un registro (tupla) y cada registro a su vez dividido en campos (atributos). La mayoría de las consultas que se realizan a una base de datos relacional (conocidas también como bases de datos tradicionales) se corresponden con *búsquedas exactas*, esto significa obtener todos los registros cuyos campos coinciden exactamente con los campos aportados durante la búsqueda. También se pueden realizar búsquedas por rango sobre valores numéricos, y búsquedas de sub-cadenas sobre campos alfabéticos; en estos casos debe existir una relación de orden sobre los campos consultados.

La información disponible en formato digital aumenta día a día su tamaño de manera ex-

ponencial. Gran parte de esta información involucra el uso de datos no estructurados tales como imágenes, sonido, texto, video, etc. Debido a que no es posible organizar estos tipos de datos en registros y campos, las tecnologías tradicionales de bases de datos para almacenamiento y búsqueda de información no son adecuadas en este ámbito.

Es en este contexto donde surgen nuevos modelos de bases de datos capaces de cubrir las necesidades de almacenamiento y búsqueda de estas aplicaciones. Nuestro interés se basa en el diseño de índices para estas nuevas bases de datos, centrándonos en bases de datos textuales y en espacios métricos.

2.0.1 Bases de Datos Textuales

Un base de datos de texto es un sistema que mantiene una colección grande de texto, y provee acceso rápido y seguro al mismo. Sin pérdida de generalidad, asumiremos que la base de datos de texto es un único texto T posiblemente almacenado en varios archivos. Las búsquedas en la que el usuario ingresa un *patrón de búsqueda* y el sistema retorna todas las posiciones del texto donde el patrón ocurre, es una de las búsquedas más comunes en este tipo de bases de datos. Si el texto es pequeño, la búsqueda de patrones puede resolverse eficientemente sin indexar el texto. Si el texto es demasiado grande se debe preprocesar el texto para construir un índice.

Mientras que en bases de datos tradicionales los índices ocupan menos espacio que el conjunto de datos indexado, en las bases de datos de texto el índice ocupa más espacio que el texto, pudiendo necesitar de 4 a 20 veces el tamaño del mismo [5, 11]. Por lo tanto construir un índice tiene sentido cuando el texto es grande, cuando las búsquedas son más frecuente que las modificaciones (de manera tal

que los costos de construcción se vean amortizados) y cuando hay suficiente espacio como para contener el índice.

Una alternativa para reducir el espacio ocupado por el índice es buscar una representación compacta del mismo, manteniendo las facilidades de navegación sobre la estructura. Pero en grandes colecciones de texto, el índice aún comprimido suele ser demasiado grande como para residir en memoria principal [6, 7]. Por esta razón, el estudio de índices comprimidos y en memoria secundaria para búsquedas en texto es un tema de creciente interés en la comunidad de bases de datos.

2.0.2 Espacios Métricos

El modelo de espacios métricos permite formalizar el concepto de búsqueda por similitud en bases de datos no tradicionales [2].

Un espacio métrico está formado por un conjunto de objetos \mathcal{X} y una función de distancia d definida entre ellos que mide cuan diferentes son. La base de datos será un subconjunto finito $\mathcal{U} \subseteq \mathcal{X}$.

Una de las consultas más comunes en este modelo de bases de datos es la *búsqueda por rango*. En esta búsqueda dado un elemento $q \in \mathcal{X}$, al que llamaremos *query* y un radio de tolerancia r , la búsqueda por rango consiste en recuperar los objetos de la base de datos cuya distancia a q no sea mayor que r . Para evitar examinar exhaustivamente la base de datos, se preprocesa la misma por medio de un *algoritmo de indexación* con el objetivo de construir una *índice*, diseñado para ahorrar cálculos en el momento de la búsqueda. En [2] se presenta un desarrollo unificador de las soluciones existentes en la temática. Básicamente se pueden distinguir dos grupos de algoritmos: *basados en pivotes* y *basados en particiones compactas*.

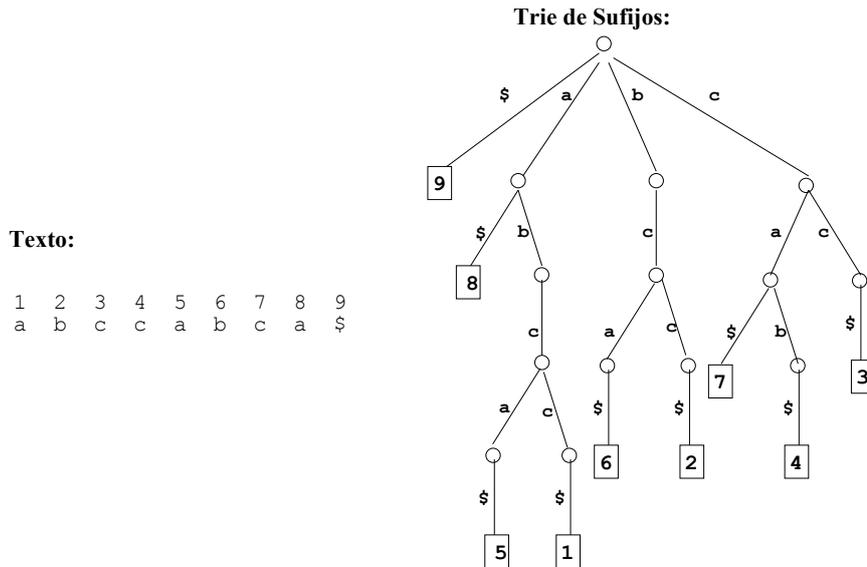


Figure 1: Un ejemplo de un texto y su correspondiente trie de sufijos.

3 Líneas de Investigación

3.1 Índices Comprimidos para Bases de Datos de Texto

Como ya mencionáramos, el principal problema que surge al indexar una bases de datos de texto es el espacio ocupado por el índice.

En bases de datos tradicionales, si construimos un índice para una relación R la cantidad de puntos de indexación está dado por la cantidad de nuplas de R y no por el espacio ocupado por R : si la relación ocupa k bytes y tiene n nuplas, en el índice existirán n puntos de indexación; notar que siempre $n < k$. En bases de datos de texto, cada caracter del texto debe ser considerado dentro del índice, en consecuencia la cantidad de puntos de indexación está dado por el tamaño del texto: si el texto ocupa k bytes, existirán k puntos de indexación.

Una forma de tratar con este problema es buscar una representación compacta del índice, manteniendo las facilidades de navegación sobre la estructura. Esto significa encontrar una representación que ocupe menos espacio que la representación clásica, pero que permita navegar sobre el índice sin necesidad

de descomprimirlo [3, 4, 6, 7, 10, 12, 13, 15].

Un *trie de sufijos* es un índice que permite resolver eficientemente las operaciones de búsquedas en texto pero que necesita en espacio 10 veces el tamaño del texto indexado. Por ejemplo, si construimos un trie de sufijos sobre un texto de 10GB, el espacio requerido para almacenarlo será de 100GB. Por esta razón, el diseño de técnicas de representación compactas para este índice es de interés.

La representación habitual de un trie consiste en mantener en cada nodo los punteros a sus hijos, junto con el rótulo correspondiente a cada uno de ellos. Existen distintas variantes de representación que consisten en organizar estos punteros a los hijos sobre una lista secuencial, sobre una lista vinculada o sobre una tabla de hashing [9]. Una de las propuestas de representación que mejor desempeño tiene en memoria principal es la de Kurtz, quien basándose en la idea de la representación sobre una lista vinculada, propuso que cada nodo mantenga un apuntador al primer hijo y almacenar los nodos hermanos en posiciones consecutivas de memoria. Esto permite durante una búsqueda, realizar una búsqueda binaria

sobre los rótulos para decidir por cual hijo seguir.

La mayoría de las propuestas existentes mantienen explícitamente la forma del árbol con punteros, los que pueden ser punteros físicos (direcciones de memoria principal) o punteros lógicos (posiciones de un arreglo).

En [14] se presenta una nueva representación de un trie de sufijos que permite reducir el espacio necesario para almacenar el índice, eliminando la necesidad de mantener los punteros explícitos a los hijos. Esta representación surge como una extensión a árboles r-arios de la técnica presentada en [8] y tiene la ventaja de permitir un posterior proceso de paginado para manejar eficientemente el trie de sufijos en memoria secundaria [16].

Notar que la información contenida en el trie está compuesta por: la forma del árbol, el rótulo de cada rama, el valor de salto de cada nodo, el grado de cada nodo y el índice del sufijo asociado a cada hoja. En [14] se propone representar de manera secuencial cada una de estas componentes, manteniendo la posibilidad de navegar eficientemente sobre el trie.

Hemos realizado una implementación que mejora en espacio a la anterior en un 40%, sin afectar los tiempos de búsqueda. Esta nueva versión compacta del trie de sufijos consiste en usar códigos *DAC* (*Directly Addressable Variable-Length Code* [1]), para los arreglos que representan la secuencia de saltos y de grados. La navegación sobre esta nueva representación sigue los lineamientos generales propuestos en [14], adaptándolo a los códigos *DAC*.

Con respecto al trabajo futuro, nos proponemos integrar esta nueva representación con la técnica de paginado propuesta en [14], a fin de lograr un índice comprimido en memoria secundaria.

3.2 Espacios Métricos y Comercio Electrónico

El comercio electrónico, también conocido como e-commerce consiste en la compra y

venta de productos o de servicios a través de la web. El desarrollo de nuevas tecnologías han permitido que la capacidad y volumen de las comunicaciones se expanda de una manera exponencial, lo que ha facilitado que el comercio electrónico tenga también un crecimiento exponencial.

Para el desarrollo de un sitio de comercio electrónico hay varios problemas que deben resolverse tales como administración de categorías de productos, búsqueda de productos, encriptación de datos, registros de usuarios, administración de medios de pago, entre otros. En este trabajo nos hemos centrado básicamente en el problema de búsqueda de productos.

Nuestro objetivo es utilizar búsquedas por similitud sobre las descripciones asociadas a los productos con el fin de estudiar el desempeño de los algoritmos basados en pivotes en un caso real de estudio.

Se ha diseñado e implementado el sistema que permite cumplir con el objetivo mencionado. Nos encontramos en la etapa de evaluación experimental del mismo a fin de establecer cuáles de las variantes utilizadas se adapta mejor a este caso de estudio.

4 Resultados Esperados

Se espera obtener índices eficientes, tanto en espacio como en tiempo, para el procesamiento de consultas en bases de datos textuales y en espacios métricos. Los mismos serán evaluados tanto analíticamente como empíricamente. Para esto último se cuenta con un conjunto de lotes de prueba usados y aceptados por la comunidad científica del área de estudio. Los mismos se encuentran disponibles en los sitios <http://pizzachili.dcc.uchile.cl> y <http://sisap.org/Home.html>.

5 Recursos Humanos

El trabajo desarrollado en esta línea forma parte del desarrollo de un Trabajo Final de la Licenciatura, una Tesis de Maestría y una Tesis de Doctorado, todas ellas en el área temática de Ciencias de la Computación, en la Universidad Nacional de San Luis.

References

- [1] Nieves R. Brisaboa, Susana Ladra, and Gonzalo Navarro. Directly addressable variable-length codes. In *SPIRE*, pages 122–130, 2009.
- [2] E. Chávez, G. Navarro, R. Baeza-Yates, and J.L. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, September 2001.
- [3] P. Ferragina and G. Manzini. Indexing compressed text. *J. ACM*, 52(4):552–581, 2005.
- [4] P. Ferragina, G. Manzini, V. Mäkinen, and G. Navarro. Compressed representations of sequences and full-text indexes. *ACM Trans. Algorithms*, 3(2):20, 2007.
- [5] G. H. Gonnet, R. Baeza-Yates, and T. Snider. *New indices for text: PAT trees and PAT arrays*, pages 66–82. Prentice Hall, New Jersey, 1992.
- [6] R. González and G. Navarro. A compressed text index on secondary memory. In *Proc. 18th International Workshop on Combinatorial Algorithms (IWOCA)*, pages 80–91. College Publications, UK, 2007.
- [7] R. González and G. Navarro. Compressed text indexes with fast locate. In *Proc. 18th Annual Symposium on Combinatorial Pattern Matching (CPM)*, LNCS 4580, pages 216–227, 2007.
- [8] N. Herrera and G. Navarro. Árboles de sufijos comprimidos en memoria secundaria. In *Proc. XXXV Latin American Conference on Informatics (CLEI)*, Pelotas, Brazil, 2009.
- [9] A. Thomo M. Barsky *, U. Stege. A survey of practical algorithms for suffix tree construction in external memory. In *Software: Practice and Experience*, 2010.
- [10] V. Mäkinen and G. Navarro. *Compressed Text Indexing*, pages 176–178. Springer, 2008.
- [11] U. Manber and G. Myers. Suffix arrays: A new method for on-line string searches. *SIAM Journal of Computing*, 22(5):935–948, 1993.
- [12] G. Navarro. Indexing text using the ziv-lempel trie. *Journal of Discrete Algorithms (JDA)*, 2(1):87–114, 2004.
- [13] G. Navarro and K. Sadakane. *Compressed Tree Representations*. Springer, 2nd edition, 2015.
- [14] D. Ruano and N. Herrera. Representación secuencial de un trie de sufijos. In *XX Congreso Argentino de Ciencias de la Computación*, Buenos Aires, Argentina, 2014.
- [15] K. Sadakane. New text indexing functionalities of the compressed suffix arrays. *J. Algorithms*, 48(2):294–313, 2003.
- [16] J. Vitter. External memory algorithms and data structures: Dealing with massive data. *ACM Computing Surveys*, 33(2):209–271, 2001.

Confluencia de Áreas de Conocimiento en un Laboratorio de Sistemas Inteligentes

Klenzi, Raúl; Araya, Jorge; López, Marcelo, Murazzo, María
Instituto de Informática / Departamento Informática / Facultad de Ciencias Exactas Físicas
y Naturales / Universidad Nacional de San Juan
Domicilio: Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas",
Rivadavia, San Juan, CPA: J5402DCS, 0264-260353 0264-4260355
{rauloscarklenzi; jorgemaraya; marcelo.sanjuan.ar; maritemurazzo}@gmail.com

Resumen

Recientemente en el ámbito del Instituto de Informática de la Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de San Juan (IdeI-FCEFNU-UNSJ) se ha conformado el “Laboratorio de Sistemas Inteligentes para la búsqueda de Conocimiento en Datos Masivos”, y está integrado por docentes investigadores pertenecientes al Departamento de Informática de la citada Facultad (DI). Las áreas de conocimiento involucradas en el laboratorio son las derivadas de dos proyectos de investigación CICITCA-UNSJ “Extracción de Conocimiento en Datos Masivos” y “Cloud Computing con herramientas libres para la evaluación de modelos de despliegue híbrido”.

Como epílogo de aquellos y prólogo de futuros proyectos en el marco del citado Laboratorio, se presenta el tratamiento de dos casos derivados de aquellas líneas de investigación. Un primer trabajo pretende automatizar el proceso de determinación de código Dewey mediante tareas de minería de texto en tanto un segundo caso propone detectar y modelar un ataque a un servidor de red por denegación de Servicios. En ambos casos las aplicaciones se han resuelto mediante la utilización de módulos y algoritmos específicos pertenecientes al entorno de software libre RapidMiner (RM) 5.3.015.

Palabras clave: Extracción de Conocimiento, TextMining, DNS, software libre

Contexto

En el ámbito del departamento Informática e Instituto de Informática DI-FCEFNU-UNSJ, IdeI-FCEFNU-UNSJ se han desarrollado en el pasado bienio dos proyectos de áreas complementarias. Por un lado el proyecto CICITCA 21/E951 “Extracción de Conocimiento en Datos Masivos” y por otro el proyecto “Cloud Computing con herramientas libres para la evaluación de modelos de despliegue híbrido” CICITCA 21/E973 y que han sido el corolario de una línea de investigación iniciada en el último lustro. Al momento y en el ámbito del IdeI, por ordenanza 02/2015-CD-FCEFNU, se acaba de conformar el laboratorio de Sistemas Inteligentes para la extracción de Conocimiento en Datos Masivos en el que confluyen ambas líneas de investigación y al que conforman docentes e investigadores de aquellos proyectos. Los casos de aplicación que se presentan, hacen las veces del prólogo de este nuevo ámbito de trabajo.

Desde el proyecto “Extracción de Datos Masivos” se han desarrollado aplicaciones, en el área temática del

Descubrimiento de Conocimiento en Bases de Datos (KnowledgeDiscovery Data -KDD-) el cual es un análisis automático exploratorio y modelado de grandes depósitos de datos. KDD es el proceso organizado de determinación de patrones válidos, nuevos, útiles y comprensibles de grandes y complejos conjuntos de datos. Minería de datos (Data Mining -DM-) es el centro del proceso de KDD, e implica la inferencia de algoritmos que exploran los datos, para desarrollar modelos y descubrir patrones previamente desconocidos. Maimon, O., & Rokach, L. (Eds.). (2005).

Estas aplicaciones han permitido concluir diferentes trabajos finales de grado así como avances importantes en otros trabajos de grado y posgrado, los cuales han contemplado las áreas de Data Mining (DM), Text Mining (TM), Web Mining (WM) y Web Analytics (WA) que mediante herramientas de software libre, en plataformas de hardware multicore, GPU computing, cluster de computadoras y sobre datos del área de la astronomía, redes sociales y relevamientos propios permitió extraer conocimiento en datos masivos.

Así mismo desde el proyecto “Cloud Computing con herramientas libres para la evaluación de modelos de despliegue híbrido” CICITCA 21/E973, entre otras actividades, se destacan la posibilidad constante de medir el flujo de bits que acceden a determinados servidores de nuestra facultad a la vez que se ha logrado poner en funcionamiento un cluster de computadoras, el cual se encuentra en una etapa de evaluación sobre el que se pretende extender las aplicaciones de KDD ya testeadas en plataformas monousuarios.

Desde la medición del flujo de bits que ingresan al servidor se propone una

primera instancia de modelación de las tramas que caracterizan a un ataque por denegación de servicios, también llamado ataque DoS (en inglés Denial of Service) o DDoS (Distributed Denial of Service), el cual es un ataque a un sistema de computadoras o red que causa que un servicio o recurso sea inaccesible a los usuarios legítimos. Normalmente provoca la pérdida de la conectividad de la red por el consumo del ancho de banda de la misma sobrecarga de los recursos computacionales del sistema de la víctima. (Mens, J. P. 2008).

Introducción

Caso de Aplicación 1) Este caso de aplicación es la continuación del trabajo presentado en el WICC 2014 “Minería de Texto en la Determinación Automática de Código Dewey (Una Primer Aproximación)” (Klenzi, R. O., & Araya, J. M.; 2014).

El objetivo de este trabajo consiste en reemplazar la tarea manual de la persona encargada de asignar la correspondiente codificación Dewey (Sistema de Clasificación Decimal creado por Melvil Dewey, un bibliotecario estadounidense, cuyo propósito inicial fue organizar la colección de la biblioteca del Amherst College.) a nuevas publicaciones bibliográfica, mediante un proceso automático basado esencialmente en similitudes sintácticas, derivadas del área de la minería de texto, entre ese nuevo material y material bibliográfico con su Dewey ya asignado.

En aquella presentación y desde una representación vectorial como valija de palabras de las fuentes de texto, se utilizó la métrica de similitud del coseno para contrastar parecidos sintácticos entre títulos e índices temático correspondiente a determinada bibliografía con las áreas

de conocimiento asociadas a los diferentes departamentos de la FCEF N así como con otros títulos e índices bibliográficos ya catalogados según su codificación Dewey y desde cuyo parecido sintáctico permitía proponer, al menos, una fracción del código correspondiente.

Para la implementación se utilizó la herramienta de software libre licencia AGPL RapidMiner 5.3.015 que es un entorno de prueba de algoritmos de aprendizaje de máquina que permite desde el modelo de datos, extraer conocimiento desde los mismos permitiendo la materialización de la totalidad de los pasos que involucran el KDD desde el preprocesamiento de datos hasta la visualización de los modelos (North, M.; 2012).

En esta oportunidad se ha agregado a aquel entorno, la medida de similitud de Okapi, presentando la fórmula directamente a través de una "bolsa de las palabras" pertenecientes a los documentos en lugar de su representación vectorial. Así el documento d_j se denota por d_j y consulta q se denota por q . anotaciones adicionales son los siguientes.

$$okapi(d_j, q) = \sum_{i \in q, d_j} \ln \frac{N - df_i + 0.5}{df_i + 0.5} \times \frac{(k_1 + 1) f_{ij}}{k_1 (1 - b + b \frac{dl_j}{avdl}) + f_{ij}} \times \frac{(k_2 + 1) f_{iq}}{k_2 + f_{iq}}$$

Figura 1: Expresión matemática de la métrica de Okapi.

Dónde:

N: Representa el número de documentos en la colección.

df_i: Representa el número de documentos que contienen al término i .

k₁, b, k₂: Son parámetros de ajuste, cuyos valores se ajustan según se obtengan los mejores resultados.

dl_j: Es el tamaño del documento en palabras.

Avdl: Representa el tamaño promedio de los documentos en la colección.

f_{ij}: Es el peso del término i en el documento j .

f_{iq}: Representa el peso del término i en la consulta q .

La fórmula de recuperación Okapi dada, se ha demostrado más eficaz que la del coseno para la recuperación de consultas cortas (Liu, B. 2007).

Al momento sólo se proponen codificaciones Dewey para material bibliográfico recientemente editado o adquirido por la biblioteca de la FCEF N afin al área de conocimiento asociada al DI-FCEF N. Por ello una primera parte de la aplicación contenida en el caso 1) y desde la implementación de la métrica de Okapi permite filtrar automáticamente, material bibliográfico ya catalogado, y con mayor similitud sintáctica con los contenidos mínimos de las asignaturas de las carreras pertenecientes al DI-FCEF N. Contra esta base, se compara el material bibliográfico de adquisición reciente rescatándose los cuatros libros de mayor similitud sintáctica y dejando a criterio del catalogador de la biblioteca la decisión final como se presenta en la Figura 2. Allí se observa, a efectos de constatar el funcionamiento del sistema, en la columna DEWEY_EXPERTO los valores asignados por el experto en catalogación de la Biblioteca de la FCEF N así como en la columna DEWEY_PROPUESTOS los valores obtenidos automáticamente. Si bien se observan diferencias, se está en el proceso de mantener nuevas reuniones con el experto a efectos de agregar al simple análisis sintáctico, reglas de análisis propias del catalogador que posibiliten una mejor aproximación a lo expresado por el experto.

Row No.	Simbolo	DEWEY	PROYECTOS	DEWEY_EXPERTO
1	COMUNICACION DE DATOS, REDES DE COMPUTADORAS Y SISTEMAS ABIERTOS	0 1920	004.6	004.6
2	COMUNICACION DE DATOS, REDES DE COMPUTADORAS Y SISTEMAS ABIERTOS	0 1958	004.67	004.6
3	COMUNICACION DE DATOS, REDES DE COMPUTADORAS Y SISTEMAS ABIERTOS	0 1721	004.6	004.6
4	COMUNICACION DE DATOS, REDES DE COMPUTADORAS Y SISTEMAS ABIERTOS	0 1675	004.68	004.6
5	GUIA DE APRENDIZAJE PHP	0 1895	005.133C	004.62
6	GUIA DE APRENDIZAJE PHP	0 1865	005.133D	004.62
7	GUIA DE APRENDIZAJE PHP	0 1417	005.11	004.62
8	GUIA DE APRENDIZAJE PHP	0 1395	005.73C	004.62
9	INTRODUCCION A LA INVESTIGACION DE OPERACIONES	0 3059	003	003
10	INTRODUCCION A LA INVESTIGACION DE OPERACIONES	0 3052	003	003
11	INTRODUCCION A LA INVESTIGACION DE OPERACIONES	0 3143	003	003
12	INTRODUCCION A LA INVESTIGACION DE OPERACIONES	0 2243	003	003
13	INVESTIGACION DE OPERACIONES	0 1497	003	003
14	INVESTIGACION DE OPERACIONES	0 1473	003	003
15	INVESTIGACION DE OPERACIONES	0 1454	003	003
16	INVESTIGACION DE OPERACIONES	0 0916	003	003
17	JAVA 2	0 4018	005.133D	005.133J
18	JAVA 2	0 3923	005.133C	005.133J
19	JAVA 2	0 3563	005.133J	005.133J
20	JAVA 2	0 3145	005.76	005.133J
21	REPARACION AVANZADA DE PC CON WINDOWS	0 1338	004.8	004.20288
22	REPARACION AVANZADA DE PC CON WINDOWS	0 1321	004.82	004.20288
23	REPARACION AVANZADA DE PC CON WINDOWS	0 1108	004.21	004.20288
24	REPARACION AVANZADA DE PC CON WINDOWS	0 1102	005.133D	004.20288

Figura 2: Propuestas de Códigos Dewey para nuevo material bibliográfico.

Caso de Aplicación 2) Este segundo caso de aplicación y desde datos representativos de una trama de bits de acceso al servidor se generó el formato csv que se aprecia en la Figura 3.

Row No.	No.	Time	Source	Destination	Protocol	Length	Info
264574	2645	78.1221590000	?	?	78	1	NPU-16, NIS-0, DSAP 088 Group, SSAP Nstar Response
264575	2645	78.1221590000	?	?	78	1	NPU-16, NIS-0, DSAP HP Extended LLC Group, SSAP Nstar Response 14PC2T
264576	2645	78.1218770000	?	?	78	1	NPU-16, NIS-0, DSAP HP Extended LLC Group, SSAP Nstar Response
264577	2645	78.1218810000	?	?	78	1	NPU-16, NIS-0, DSAP Ungermann-Bass Individual, SSAP Nstar Response
264578	2645	78.1218850000	?	?	78	1	NPU-16, NIS-0, DSAP Ungermann-Bass Group, SSAP Nstar Response
264579	2645	78.1218890000	?	?	78	1	Unknown Type
264580	2645	78.1218930000	?	?	78	1	NPU-16, NIS-0, DSAP Rounde Program Load Group, SSAP Nstar Response
264581	2645	78.1218970000	?	?	78	1	Unknown PC protocol 003
264582	2645	78.1219170000	192.168.1.36	192.168.1.37	TCP	54	0 - 36990 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264583	2645	78.1227730000	192.168.1.36	192.168.1.37	TCP	54	0 - 36991 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264584	2645	78.1225250000	?	?	78	1	NPU-16, NIS-0, DSAP ESC Network Layer Group, SSAP Nstar Response
264585	2645	78.1225290000	?	?	78	1	NPU-16, NIS-0, DSAP NULL LAMP Individual, SSAP 088 Command
264586	2645	78.1226290000	?	?	78	1	NPU-16, NIS-0, DSAP NULL LAMP Group, SSAP 088 Command
264587	2645	78.1261000000	192.168.1.36	192.168.1.37	TCP	54	0 - 36992 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264588	2645	78.1268310000	192.168.1.36	192.168.1.37	TCP	54	0 - 36993 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264589	2645	78.1264650000	192.168.1.36	192.168.1.37	TCP	54	0 - 36994 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264590	2645	78.1266550000	192.168.1.36	192.168.1.37	TCP	54	0 - 36995 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264591	2645	78.1268510000	192.168.1.36	192.168.1.37	TCP	54	0 - 36996 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264592	2645	78.1271250000	192.168.1.36	192.168.1.37	TCP	54	0 - 36997 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264593	2645	78.1275900000	192.168.1.36	192.168.1.37	TCP	54	0 - 36998 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264594	2645	78.1287900000	192.168.1.36	192.168.1.37	TCP	54	0 - 36999 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264595	2645	78.1287770000	192.168.1.36	192.168.1.37	TCP	54	0 - 37000 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264596	2645	78.1291450000	192.168.1.36	192.168.1.37	TCP	54	0 - 37001 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264597	2645	78.1291800000	192.168.1.36	192.168.1.37	TCP	54	0 - 37002 RST, ACK Seq=1 Ack=1 Win=0 Len=0
264598	2645	78.1291100000	?	?	78	1	NPU-16, NIS-0, DSAP LLC 802-Layer Management Individual, SSAP 088 Command

Figura 3: Trama de datos de entrada

Esta fuente con más de 800.000 registros, sirvió para estudiar y modelar las condiciones que evidencian un ataque por denegación de Servicios. La misma y tras tareas de preprocesamiento donde se filtraron registros con valores faltantes, se transformó el atributo info mediante tareas de filtrado de texto, se estableció una ventana temporal de análisis y desde la aplicación de un clasificador y en RM 5.3.015, se logró modelar lo que en principio permite reconocer un ataque por denegación de servicios.

La aplicación inicialmente emula el análisis de un experto en el análisis de ataques por denegación de servicios y posteriormente modela la salida “label” “flood” bivaluada asignando un 1 al reconocimiento de un ataque y 0 en caso contrario. Sobre esta nueva columna se

realiza un clasificador basado en árboles que modela el ataque por denegación de servicios y cuya validación se realiza mediante el módulo específico del entorno de software RM logrando una performance expuesta en la Figura 4.

Confusion Matrix			
	true false	true true	class precision
pred false	121500	0	100.00%
pred true	2	194773	100.00%
class recall	100.00%	100.00%	

Figura 4: Matriz de confusión que presenta las predicciones de un ataque por denegación de servicios

Líneas de Investigación, Desarrollo e Innovación

El objetivo del laboratorio recientemente creado es profundizar en las tareas de investigación aplicada en lo concerniente a extracción de conocimiento en datos masivos para lo que se deberá propender a la utilización de hardware distribuido o a nivel del cloud computing amalgamando los conocimientos y actividades de cada uno de los grupos de investigación de los que deriva el laboratorio. En particular los datos masivos habrán de ser recabados y provistos por el Grupo de Astronomía Extragaláctica (GAE) de la FCFEN sobre los que habrán de ejecutarse algoritmos de extracción de conocimiento basados en entornos de aprendizaje de máquina que se ejecuten en plataformas paralelo-distribuidas.

Resultados y Objetivos

Los resultados que marcan el prólogo de la formación del laboratorio resultan auspiciosos como lo evidencian los trabajos presentados, y áreas temáticas comunes en que cada uno de los grupos

de investigación original hará de soporte al otro, generando una sinergia favorable. Según la ordenanza, 02/2015-CD-FCEF, de creación del laboratorio de sistemas inteligentes para la búsqueda de conocimiento en datos masivos, el mismo tiene como objetivo, extraer conocimiento desde grandes bases de datos mediante la utilización de algoritmos de minería de datos y aprendizaje de máquina soportado por arquitecturas secuenciales y paralelas-distribuidas

Se espera obtener conocimiento, sobre datos del área de la astronomía, que resulten ser fiable y contrastable con aplicaciones estadísticas llevadas adelante por el grupo GAE, pero que desde aplicaciones distribuidas y herramientas de software aprendizaje de máquina y extracción de conocimiento, permitan mejorar la performance de resultados anteriores.

Formación de Recursos Humanos

En el último año desde la actividad de ambos proyectos de investigación se han realizado las siguientes aplicaciones y trabajos de grado y posgrado que han permitido la formación de los siguientes recursos humanos.

En el transcurso del año 2015 desde la defensa de trabajos finales de grado que permitieron coronar las carreras en Licenciaturas en Ciencias de la Computación y Licenciatura en Sistemas de Información, se defendieron cinco trabajos finales enmarcados en las áreas llevadas adelante por el grupo de investigación en tareas de DM, TM, WA, mediante la aplicación a diferentes tipos de datos, de algoritmos secuenciales y paralelos, mediante la utilización de herramientas de software libre y en hardware multicore y GPU computing.

Al momento y en ambas carreras se están dirigiendo otros seis trabajos finales que seguramente habrán de ser defendidos a la brevedad.

En el ámbito de posgrado se llevan adelante cuatro tesis de maestría, así como tres integrantes del grupo han comenzado a delinear sus líneas de trabajo de Doctorado en Ciencias de la Informática.

Por último y como parte de los objetivos planteados en el proyecto 21/E951 y en el ámbito de la FCEF, se dictó el curso de posgrado “Búsqueda de Conocimiento en Datos” que con una duración de 50hs se dictó en Agosto y Setiembre de 2015, a la vez que se pretende durante el presente ciclo lectivo brindar un curso en el marco de la misma temática pero con profundidad.

Referencias

- Klenzi, R. O., & Araya, J. M. (2014, October). Minería de texto en la determinación automática de código Dewey. XVI Workshop de Investigadores en Ciencias de la Computación.
- Liu, B. (2007). Web data mining: exploring hyperlinks, contents, and usage data. Springer Science & Business Media.
- Maimon, O., & Rokach, L. (Eds.). (2005). Data mining and knowledge discovery handbook (Vol. 2). New York: Springer.
- Mens, J. P. (2008). Alternative DNS Servers: Choice and Deployment, and Optional SQL/LDAP Back-Ends. UIT Cambridge Ltd..
- North, M. (2012). Data mining for the masses.

Evaluación de técnicas de Data Mining para la obtención de perfiles de ingresantes a la UNNOBA

Monica Sarobe, Ana Smail, Maria Rosana Piergallini, Mercedes Guasch, Marina Rodriguez, María Lanzillota, Mariana Adó, Claudia Banchoff

Instituto de Investigación y Transferencia en Tecnología (ITT), Escuela de Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)

Sarmiento y Newbery (CP 6000), Junín, Buenos Aires, Argentina. Teléfonos (0236) 4636945/44

{monica.sarobe, ana.smail, rosana.piergallini, mercedes.guasch, marina.rodriguez, maria.lanzillota, mariana.ado, cbanchoff}@itt.unnoba.edu.ar

Resumen

El ingreso a la Universidad es siempre un desafío para el aspirante como para la institución que lo albergará.

Durante la inscripción a las propuestas académicas de grado y posterior desarrollo del curso de ingreso la Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA) recolecta información a través de sus sistemas de gestión. Esta información representa un importante activo para la Universidad, en la medida en que pueda ser transformada en conocimiento ya sea para la toma de decisiones que conciernen a la población estudiantil, como para el diseño de planes de acción.

La presente línea de investigación propone procesar los datos recolectados por los sistemas de gestión de la UNNOBA durante el ingreso, para obtener en principios patrones o semejanzas entre los aspirantes con idénticos resultados en el curso de ingreso. Esto permitirá entre otras cosas

determinar estrategias de aprendizajes adaptativos.

Palabras clave: KDD, Data Mining

Contexto

Esta línea de investigación forma parte del proyecto “Herramientas didácticas y nuevas formas de interacción para la inclusión de TICs en la enseñanza” aprobado por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA en el marco de la convocatoria a Subsidios de Investigación BIANUALES (SIB2015).

El proyecto se desarrolla en el Instituto de Investigación en Tecnologías y Transferencia (ITT) dependiente de la mencionada Secretaría, y se trabaja en conjunto con la Escuela de Tecnología de la UNNOBA.

Introducción

La Universidad Nacional del Noroeste de la Provincia de Buenos Aires

(UNNOBA) fue creada en el año 2004, comenzando a dictar sus carreras de grado en el año 2005. Este mismo año implementó por primera vez el dictado de un curso de ingreso de modalidad presencial. Luego, a partir del año 2010 se sumó la modalidad semipresencial; y posteriormente, a partir del año 2013, se incorpora la opción de rendir una evaluación de conocimientos. Dicha evaluación, permite a los aspirantes a carreras validar sus conocimientos mediante la aprobación de un examen, quedando así exceptuados realizar el curso de ingreso.

Los aspirantes expresan la voluntad de ingresar a la Universidad realizando una preinscripción a carreras en el sistema SIU-Preinscripción. Este sistema recolecta datos demográficos, sociales, económicos y familiares en una base de datos relacional. Para aquellos alumnos que no se encuentran exceptuados del curso de ingreso, se cuenta con los resultados obtenidos en las dos materias que lo componen en cada una de las modalidades (presencial, semipresencial o evaluación de conocimientos) también en una base de datos relacional.

Los datos recolectados durante el proceso de ingreso en todos estos años conforman hoy un activo de potencial relevancia dado que posibilitan obtener patrones de los ingresante a fin de poder establecer un perfil de los mismos.

La comunidad educativa de la universidad considera de real importancia la posibilidad de contar con una descripción o caracterización de los ingresantes a través de perfiles, dado que un mayor conocimiento de los mismos puede influir positivamente en la toma de decisiones concernientes al ingreso como también en el diseño de estrategias de enseñanza y aprendizaje.

Establecer perfiles es una actividad muy difundida en muchas áreas

disciplinarias, es análoga al proceso de determinar y clasificar patrones. El proceso de Extracción de Conocimientos a partir de datos (KDD, Knowledge Discovery from Databases) es un proceso no trivial para identificar patrones válidos, novedosos y potencialmente útiles a partir de los datos. Este busca el descubrimiento de conocimiento sin una hipótesis preconcebida. Involucra la preparación de los datos, su análisis o minería de datos e interpretación de los resultados. [6]

La Minería de Datos (Data Mining), involucra técnicas necesarias para la construcción de modelos a partir de la información disponible. Este proceso puede utilizar dos enfoques bien definidos: descriptivo o predictivo. En el primer enfoque, se intenta caracterizar las situaciones de manera tal de poder entender cuáles son las variables principales que describen el comportamiento de una situación particular. El segundo enfoque implica la utilización del modelo para establecer de antemano una situación problemática. [3]

En el caso de la determinación de perfiles de los ingresantes, un modelo descriptivo puede explicar las variables que mejor caracterizan a los aspirantes a carreras de grado con idénticos resultados en el curso de ingreso y obtener una visión global de los mismos. La simple aplicación de una técnica de MD a una muestra representativa de los datos pre-procesada (vista minable), y el conocimiento previo del problema, no garantizan patrones expresivos, novedosos y útiles. [2] Por lo que parece tener sentido la evaluación de diferentes técnicas o algoritmos para obtenerlos.

Líneas de Investigación, Desarrollo e Innovación

La presente investigación se encuadra dentro del eje “Tratamiento masivo de

datos”. En este sentido se pretende procesar los datos recolectados de los sistemas de gestión de la Universidad durante el ingreso, para obtener en principios patrones o semejanzas entre los aspirantes con idénticos resultados en el curso de ingreso.

La investigación actual sigue las siguientes líneas:

- Evaluar técnicas de selección de características.
- Evaluar técnicas de para obtener modelos descriptivos.
- Evaluar el ensamble de algoritmos para potenciar el carácter descriptivo de los modelos.
- Evaluar técnicas de visualización de los resultados.

Resultados y Objetivos

Durante la primera etapa del proyecto, como parte del pre-procesamiento de datos, hemos obtenido una vista minable de los datos extraídos del Sistema de Preinscripción a la Universidad, el cual cuenta con datos demográficos, económicos, sociales, entre otros; junto con los resultados de las tres modalidades en que se dicta el curso de ingreso y su resultado final.

Las tareas a realizar en la segunda etapa se centraran en:

- Reducción de la dimensión del problema, mediante la selección de los atributos más representativos de la información disponible.
- Selección del algoritmo que produzca los patrones más expresivos.
- Validación de la solución obtenida.
- Ensamblar con algoritmos que describan y expliquen mejor el conocimiento obtenido.

Se espera que la presente línea de investigación contribuya aportar información implícita, previamente desconocida y potencialmente útil para la toma de decisiones que conciernen a la población estudiantil y para el diseño de planes de acción tanto preventivos como correctivos.

Formación de Recursos Humanos

Se cuenta con dos alumnos avanzados de la carrera Licenciatura en Informática que poseen becas de iniciación a la investigación otorgadas por la UNNOBA y se encuentra en desarrollo dos tesis de posgrado en la UBA.

Bibliografía

- [1] WICC 2014 - Detección Temprana de la Deserción en Cursos Virtuales - Russo, Sarobe, Lencina, López Gil, Carrattini, Smail, Piergallini, Guash, Alonso, Ramón
- [2] Evaluación de técnicas de Extracción de Conocimiento en Bases de Datos y su aplicación a la deserción de alumnos universitarios – Formia, Lanzarini - 2012
- [3] Introduction to Data Mining – Tan, Steinbach, Kumar - 2006
- [4] Selección de atributos representativos del avance académico de los alumnos universitarios usando técnicas de visualización. Un caso de estudio – Lanzarini, Charnelli, Baldino, Diaz – 2015
- [5] Educational Data Mining by Using Neural Network – Upadhyay, Lucknow - 2016
- [6] Introducción a la Minería de Datos - Orallo, Quintana, Ramirez - 2004
- [7] Minería de datos para un Sistema de alerta temprana de deserción en carreras de Ingeniería – Ahumada, Dip, Herrera, Leguizamón, Almendra - 2015

Algoritmos para el tratamiento de imágenes de campos de luz

Gustavo Bizai, Facundo Peiretti, Adrián Salvatelli, Bartolomé Drozdowicz,
Alejandro Hadad, Diego Evin

Laboratorio de Sistemas de Información, Facultad de Ingeniería, Universidad
Nacional de Entre Ríos
Ruta 11 Km.10 Oro Verde, Entre Ríos, Argentina
E-mail: {gbizai, asalvatelli, bdrozdo}@bioingenieria.edu.ar

Resumen

El Proyecto de Investigación y Desarrollo “Sistema de Información Plenóptica como medio diagnóstico para Lámparas de Hendidura” propone la utilización de campos de luz como estrategia para agregar información tridimensional a la imagen del fondo ocular. Una de las líneas de trabajo consiste en el desarrollo de algoritmos para la obtención de enfoque dinámico, multiperspectiva y mapa de profundidades a partir de imágenes plenópticas que se generan en una única toma. La percepción de profundidades constituye un valioso apoyo a la toma de decisiones en múltiples aplicaciones. Se presentan algunas metodologías para el tratamiento computacional de capturas con una cámara plenóptica, utilizando imágenes de dominio público mientras se está trabajando en la implementación de la óptica asociada a una lámpara de hendidura. Se muestra resultados de los algoritmos y se discute su eficiencia en términos de su complejidad y tiempo de procesamiento. El trabajo futuro es optimizar estas propuestas, implementar nuevas y obtener mapas de profundidad.

Palabras clave: Campos de Luz, Fotografía Plenóptica, Pila Focal.

Contexto

La línea de I+D+i presentada es parte del Proyecto de Investigación y Desarrollo PIDUNER N°6146 “Sistema de Información Plenóptica como medio diagnóstico para Lámparas de Hendidura”, financiado por la Universidad Nacional de Entre Ríos y ejecutado en el ámbito de la Facultad de Ingeniería (FI-UNER).

Introducción

La fotografía de fondo de ojos es de especial interés para la detección temprana de patologías [1-3] así como también el seguimiento de la evolución de los tratamientos [4]. Una de las complicaciones de la diabetes es el edema macular, cuyo diagnóstico temprano es fundamental. Su evaluación clínica es siempre difícil, siendo la fotografía estéreo y la biomicroscopía con lámpara de hendidura las técnicas más utilizadas para mostrar cambios en el volumen retiniano. Sin embargo, ambas técnicas son dependientes del observador y no permiten la medición objetiva del cambio [5]. Asimismo, la lámpara de hendidura, equipo presente en cualquier consultorio oftalmológico, no ha evolucionado significativamente desde el punto de vista tecnológico [6]. La posibilidad de obtener imágenes de la topografía del fondo

ocular es de gran interés para el médico oftalmólogo como soporte objetivo para la toma de decisiones, máxime si se utiliza un equipo económico y de difusión masiva como lo es dicha lámpara. El valor agregado de información 3D de estas imágenes, incrementa su valor diagnóstico.

El PIDUNER abarca tres líneas de trabajo: a) caracterización de la lámpara de hendidura y obtención de datos para la decisión de compra de equipamiento de fotografía plenóptica; b) desarrollo de la algorítmica para la obtención de imágenes multifoco, multiperspectiva y mapa de profundidades; c) aplicación de metodologías para extracción de características de imágenes 3D. Las primeras dos líneas mencionadas se trabajan en paralelo y mientras se está desarrollando la óptica de captura, se ha comenzado a estudiar los algoritmos con imágenes plenópticas públicas dado que las metodologías son las mismas.

Este trabajo presenta la generación y evaluación de diferentes algoritmos para el renderizado de imágenes a partir de los campos de luz [7]. El campo de luz de una escena es el conjunto de todos los rayos de luz que refleja, en cualquier punto del espacio, en cualquier dirección. En otras palabras, es la radiancia en el espacio como función de la posición y dirección. Al conjunto de todos los rayos de luz que ocupa el espacio durante todo el tiempo se los describe con la función plenóptica [8], La insistencia de la “totalidad” de esta función es la razón de su nombre: óptica plena, completa, llena. La función plenóptica describe la luz en 7 dimensiones: espacio (3D), dirección (2D), frecuencia (1D) y tiempo (1D). Es posible reducir esta alta dimensionalidad a 4D si se utiliza una parametrización de dos planos paralelos [9] y se consideran imágenes estáticas. Con esta información 4D es posible obtener imágenes 2D

enfocadas en diferentes profundidades de la escena, así como también mostrar la imagen desde diferentes ángulos. Con este conjunto de imágenes denominado “pila focal”, puede reconstruirse un modelo de la topografía de la escena.

En este trabajo se presentan algunos algoritmos para renderizar imágenes en diferentes profundidades y perspectivas, a partir de los campos de luz.

Líneas de Investigación, Desarrollo e Innovación

La investigación pertenece al Área Computación Gráfica, Imágenes y Visualización. La Fotografía Plenóptica [10] es una forma de Fotografía Computacional, en la cual la cámara plenóptica [11] recolecta información más rica que una cámara convencional para luego someterla a procesos computacionales que dan como resultado una visualización de la escena en diferentes profundidades y perspectivas.

Resultados y Objetivos

La captura del campo de luz se realiza con una cámara a cuyo sensor de imagen se antepone una matriz de lenticulas.

Codificación del Campo de Luz

Las imágenes crudas consisten en un conjunto de $M \times N$ subimágenes de $P \times Q$ pixeles cada una (figura 1). $M \times N$ es la cantidad de lenticulas que constituyen la matriz de microlentes, mientras que $P \times Q$ son los pixeles del elemento sensor de la cámara, que constituyen las subimágenes de cada lenticula. El tamaño de la imagen cruda de campo de luz es de $(M \times P) \times (N \times Q)$ pixeles. A partir de estas capturas es posible una visualización multiperspectiva y refocalización

computacional. La información de cada pixel se codifica en una matriz 5D que permite direccionarlo según la subimagen a la que pertenece (posición de la lenticula M,N) y su posición dentro de ella (P,Q) , para cada canal de color (ch) . El script consiste en varios bucles anidados que distribuyen cada pixel a una posición (M,N) y una dirección (P,Q) para cada canal de color. Requiere como variables la imagen a ser decodificada, la cantidad de lenticulas y la cantidad de pixeles por lenticula y devuelve la matriz $LF(M,N,P,Q,ch)$.

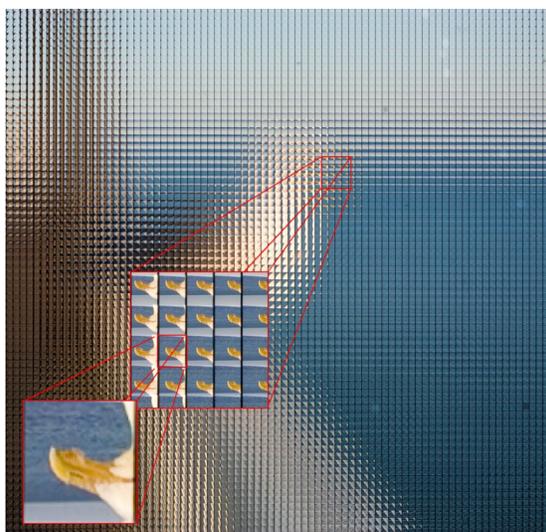


Figura 1. Fotografía Plenóptica obtenida por Georgiev et al. [11].

Para obtener una imagen en determinado ángulo, se selecciona una misma posición (P,Q) para todas las lenticulas y se genera una imagen de $M \times N$ pixeles. Recuérdese que la posición de los pixeles debajo de cada microlente guarda la información de la dirección de los rayos. En la figura 2 se observa una secuencia de imágenes en diferentes perspectivas, reconstruidas a partir de una única captura.

Filtro Desplazamiento-Suma

Uno de las estrategias más sencillas para obtener foco en diferentes profundidades es aplicando el Filtro Desplazamiento-Suma. El cómputo de refocalización consiste en desplazar cada imagen en perspectiva por un factor de $(P(1-1/\alpha), Q(1-1/\alpha))$ y sumarlas. Cuando $\alpha=1$ significa que no hay refocalización y el desplazamiento es nulo para el conjunto de las imágenes de perspectiva. En la figura 3 se observa una secuencia de imágenes enfocadas a diferentes profundidades, según el parámetro alfa.

Las imágenes refocalizadas presentan borrono, una manifestación del aliasing asociado con un resampleo inapropiado del campo de luz dado que en el cómputo de los desplazamientos los índices de posición calculados son reales y deben redondearse al entero más cercano.

Fast Discrete Focal Stack Transform

Otra aproximación a la refocalización digital es aplicar un caso particular de la Transformada de Radón y su extensión N -dimensional. En [12] se propone una Transformada para generar la pila focal, basada en la Transformada Discreta de Radon. Sus autores proponen el algoritmo FDFST (Fast Discrete Focal Stack Transform) para obtener un conjunto de imágenes enfocadas en diferentes profundidades. La figura 4 muestra el resultado de aplicar FDFST a una imagen. Se observa resultado similar al obtenido con el algoritmo de desplazamiento-suma, excepto por un nivel de brillo menor.

Filtro Planar 4D de dominio frecuencial

Otra alternativa de filtrado es trabajar en el dominio de Fourier. El aporte de Ng [13] fue la generalización del Teorema del Corte, que permite obtener imágenes refocalizadas al aplicar la transformada inversa de Fourier a un plano cortado del hiperespacio transformado del campo de

luz. Dicho plano se obtiene construyendo un filtro frecuencial H que toma como parámetro una pendiente (correspondiente a la inclinación del plano de corte). Asociada a dicha pendiente está la profundidad a la que se obtendrá el foco en la imagen resultado. Se multiplica punto a punto el espacio transformado por el filtro H y se aplica la transformada inversa de Fourier para generar la imagen mencionada. En la figura 5 se muestran algunas imágenes refocalizadas por corte del hiperespacio transformado.

Trabajo futuro

Actualmente se están implementando y evaluando otros algoritmos con el propósito de encontrar aquel que se ajuste a los requerimientos del Proyecto. Una vez seleccionado, deberá construirse la pila focal e implementarse el algoritmo para crear el mapa de profundidades del fondo ocular. La imagen del fondo visualizada como superficie 3D se espera que incremente el valor diagnóstico de los estudios retinográficos.

Formación de Recursos Humanos

El equipo de trabajo está conformado por especialistas del área de sistemas de información, inteligencia artificial y bioingeniería. Relacionadas con el Proyecto de Investigación se han generado dos Tesinas de Grado (una aprobada y la otra en curso) y una de Maestría (en curso).

Referencias

[1] Salvatelli A, Bizai G, Martina M, Hadad A and Drozdowicz B 2013 Detección temprana de patologías diabéticas oculares en retinografías utilizando un Sistema Multiagentes 42 *JAIHO - CAIS 2013* - ISSN: 1853-1881 p 263

[2] Salvatelli A, Bizai G, Barbosa G, Drozdowicz B and Delrieux C 2007 A comparative analysis of pre-processing techniques in color retinal images *Journal of Physics: Conference Series* **90** 012069. IOP Publishing. doi:10.1088/1742-6596/90/1/012069

[3] Insfran M, Yancovich J, Salvatelli A, Bizai G and Drozdowicz B 2015 Workflow based on pixel processing for automatic segmentation of vasculature in retinographies *IFMBE Proc* **49** ISBN/ ISSN 1680-0737

[4] Gelroth Balbi A, Rodriguez D, Salvatelli A, Drozdowicz B and Bizai G 2011 Analysis and implementation of methodologies for the monitoring of changes in eye fundus images *Journal of Physics: Conference Series* **332** 012036 IOP Publishing.

[5] Cunha-Vaz J 1998 Diabetic macular edema *European Journal of Ophthalm.* **8** n.3 pp. 127–130

[6] González G, Faisano A and Torres R 2015 The slit-lamp survey in Entre Ríos, Argentina *IFMBE Proc* **49**, pp 117–119.

[7] Gershun A 1939 The light field. Moscow, 1936. *Journal of Mathematics and Physics* **XVIII**, 51–151 Translated by P. Moon and G. Timoshenko.

[8] Adelson E and Bergen J 1991 The plenoptic function and the elements of early vision *Computational models of visual processing* **91** n.1 pp 3–20

[9] Levoy M and Hanrahan P 1996 Light field rendering *SIGGRAPH ACM* pp 31–42

[10] Adelson E and Wang J 1992 Single lens stereo with a Plenoptic Camera *IEEE transactions on pattern analysis and machine intelligence* **14** n.2

[11] Georgiev, T.: New results on the Plenoptic 2.0 camera. *Signals, Systems and Computers, Conference Record of the Forty-Third Asilomar Conference*. Pp 1243–1247 (2009)

[12] Pérez Nava F, Marichal-Hernández J and Rodríguez-Ramos J 2008 The discrete Focal Stack transform *Proceedings of Eusipco'08*, August 25–29 2008

[13] Ng R 2006 Digital light field photography *Ph.D. dissertation, Stanford University*



Figura 2: Dos perspectivas de la misma escena. En el círculo, detalle de una de las diferencias.



Figura 3: Resultados obtenidos del algoritmo Shift Sum para distintos valores de 'alfa'.

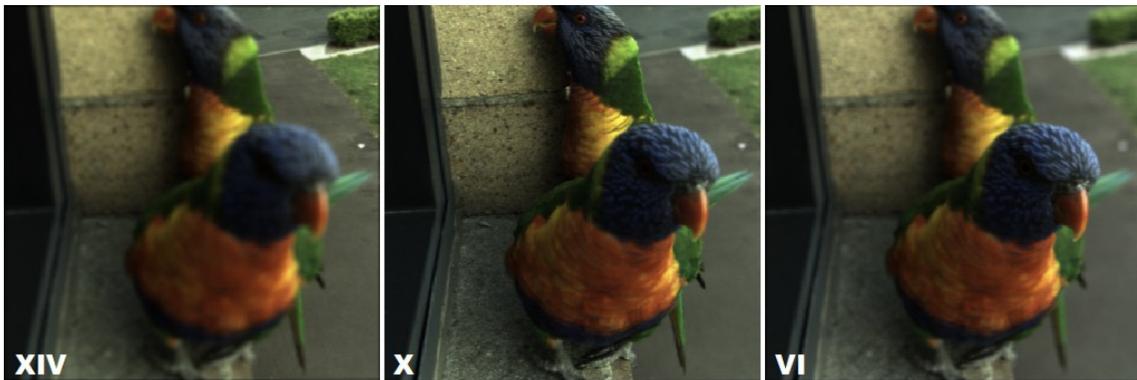


Figura 4: Tres de las 17 imágenes de una pila focal.

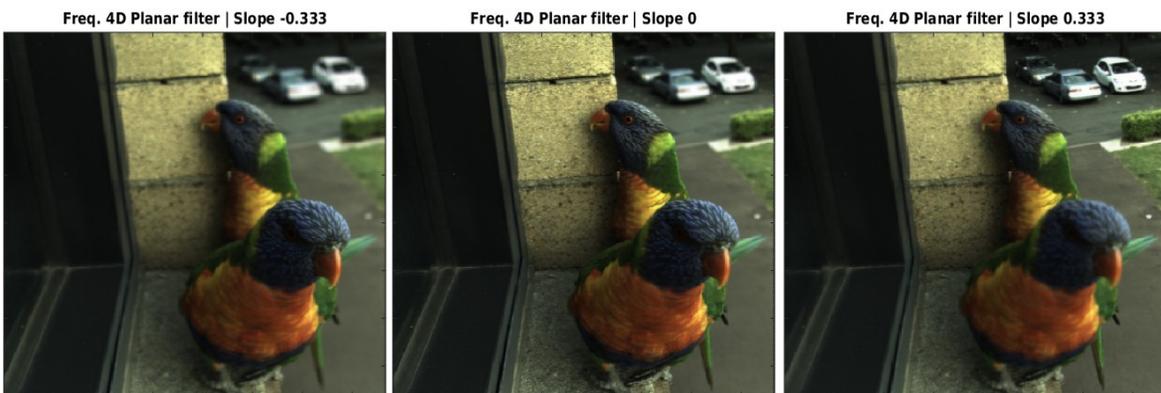


Figura 5: Filtro planar frecuencial 4D con diferentes pendientes.

Generación de Características y Reconocimiento Estadístico de Patrones

Luciano Lorenti, Lucía Violini, Javier Giacomantone, María José Abásolo,
Oscar Bria, Marcelo Naiouf, Armando De Giusti

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – UNLP

{llorenti, lviolini, jog, mjabasolo, obria, mnaiouf, degiusti}@lidi.info.unlp.edu.ar

CONTEXTO

Esta línea de investigación y desarrollo (I/D) forma parte del proyecto “Cómputo Paralelo de Altas Prestaciones (HPC). Fundamentos y Evaluación de Rendimiento en HPC. Aplicaciones a Sistemas Inteligentes, Simulación y Tratamiento de Imágenes – 11/F017”, del Instituto de Investigación en Informática LIDI acreditado por la UNLP en el marco del Programa de Incentivos.

RESUMEN

Este trabajo describe una línea de I/D y los resultados esperados de la misma. El objetivo principal es analizar, desarrollar y evaluar modelos y métodos computacionales. A partir de los resultados obtenidos y los métodos propuestos el segundo objetivo es la transferencia de los mismos y el estudio de las técnicas de enseñanza-aprendizaje más adecuadas para los temas abordados.

Los dos ejes principales de investigación son, generación de características y reconocimiento estadístico de patrones. Se analizan métodos de generación de características a partir de señales en general y de imágenes digitales en particular, considerando el método de aprendizaje supervisado subyacente y su poder de discriminación. El segundo eje se centra en estudiar, desarrollar y evaluar métodos computacionales de reconocimiento estadístico de patrones, en particular métodos de clasificación supervi-

sada, no supervisada y reducción de dimensión. Todos los sistemas de reconocimiento de patrones diseñados o analizados responden a un modelo integral, en los cuales el énfasis está dado en el análisis científico de cada etapa.

Palabras Clave: Reconocimiento de Patrones. Análisis de Imágenes. Aprendizaje Automático. Generación de Características.

1. INTRODUCCION

En general todo tipo de fenómeno físico o químico del cual podamos directa o indirectamente extraer descriptores viables, es potencialmente tratable por un sistema automático de reconocimiento de patrones. Algunos de ellos tienen protocolos precisos para su implementación producto de años de investigación básica. Es necesario por lo tanto diferenciar claramente las hipótesis planteadas para proponer modelos o métodos. Es de particular interés en esta línea de I/D el estudio de la adecuada generación de características a partir de imágenes digitales. El trabajo actual lo podemos clasificar en cuatro tópicos principales bien diferenciados correspondientes a sub-disciplinas dentro del área. La primera sub-disciplina es clasificación supervisada donde el énfasis de nuestro trabajo se centra en el estudio de métodos de clasificación basados en núcleos dispersos, en particular máquinas de soporte vectorial. La segunda es clasificación no supervisada donde la principal línea

de trabajo son las técnicas de agrupamiento. Actualmente con énfasis en detección de valores atípicos, métodos basados en teoría espectral de grafos y su aplicación a imágenes de rango. El tercer tópico de fundamental importancia es el de reducción de dimensión en particular selección de características. El cuarto tópico que involucra a los tres anteriores puede ser denominado como de transferencia y técnicas de enseñanza-aprendizaje específicas para los temas de esta línea de investigación.

2. LINEAS DE INVESTIGACION y DESARROLLO

- Selección y extracción de características.
- Máquinas de soporte vectorial. Kernels y algoritmos de optimización.
- Clasificación no supervisada. Técnicas de agrupamiento (clustering).
- Detección de anomalías.
- Modelos y métodos computacionales en Análisis de Imágenes y Reconocimiento de Patrones.
- Métodos de estimación de parámetros para clasificadores Bayesianos.
- Métricas y pseudométricas.
- Criterios de evaluación de desempeño en sistemas de clasificación automática.
- Criterios y algoritmos para combinación de clasificadores.
- Clasificación de series temporales y clasificación contextual.
- Análisis de Señales Digitales.

3. RESULTADOS OBTENIDOS /ESPERADOS

- Desarrollar modelos y optimizar algoritmos particulares de clasificación supervisada y no supervisada.
- Evaluación de los métodos de análisis de desempeño y su aplicación sobre los clasificadores y conjuntos de datos propuestos.

- Construcción de una mesa multi-táctil basada en visión por computador para su uso en educación especial [29].
- Se estudiaron y propusieron métodos para detección en series temporales de fMRI [23][32].
- Se realizó la evaluación de rendimiento en sistemas de reconocimiento de patrones supervisados y de clasificación binaria [16].
- Se desarrollaron métodos de segmentación de imágenes de rango [20][30][33].
- Desarrollo de técnicas de selección en espacios multi-dimensionales [31].

4. FORMACION DE RECURSOS HUMANOS

La formación de recursos humanos en esta línea I/D se puede clasificar como indirecta o directa. En forma directa, actualmente, hay dos investigadores realizando su doctorado y cuatro alumnos han concluido tesis. Indirectamente como área permanente de consulta y transferencia para investigadores de líneas de I/D estrechamente relacionadas.

6. BIBLIOGRAFIA

1. Batagelj V, Bock H, Ferligoj A. "Data Science and Classification". Springer, (2006).
2. Fukunaga K. "Introduction to Statistical Pattern Recognition". Second Edition. Academic Press, (1990).
3. Devijer P, Kittler, J. "Pattern Recognition: theory and applications". Springer, (1986).
4. Anke Meyer-Base. "Pattern Recognition for Medical Imaging". Academic Press, (2004).
5. Kim H.Y., Giacomantone J. O., Cho, Z. H. Robust Anisotropic Diffusion to Produce Enhanced Statistical Parametric Map, Computer Vision and Image Understanding, v.99, pp.435-452 (2005).
6. Kim H.Y., Giacomantone J. O., A New Technique to Obtain Clear Statistical Para-

- metric Map by Applying Anisotropic Diffusion to fMRI, IEEE, International Conference on Image Processing. Proceedings, Genova, Italy, v.3, pp.724-727 (2005).
7. Corte C, Vapnik V, Support vector networks. Machine Learning v.20, pp.273-297 (1995).
 8. Vapnik, V. The Nature of Statistical Learning Theory. N. Y. Springer (1995).
 9. Li S., Fevens L., Krzyzak A., Li S. Automatic Clinical Image Segmentation Using Pathological Modelling, PCA and SVM, MLDN, LNAI 3587 pp.314-324, (2005).
 10. Lei Z., Yang Y., Wu Z.. Ensembles of Support Vector Machine for Text-Independent Speaker Recognition, IJCSNS v.6 n.5A pp. 163-167, (2006).
 11. Boekhorst R., Abnizova I., Wernich L. Discrimination of regulatory DNA by SVM on the basis of over and under-represented motifs, ESANN pp. 481-486 (2008).
 12. Vossen Anselm. Support Vector Machines in High Energy Physics, CERN, Geneva, Switzerland, pp.23-33 (2005).
 13. Wang L., Chang M., Feng J. Parallel and Sequential Support Vector Machines for Multi-label Classification, International Journal of Information Technology, v.11 n.9 pp. 11-18, (2005).
 14. Rüping S., SVM kernels for time series analysis, G1-Workshop-Woche Lernen-Lehren-Wissen-Adaptivitet, pp.43-50 (2001).
 15. Yang K., Shahabi C. A pca-based kernel for kernel pca on multivariate time series, IEEE Intern. Conf. on Data Mining (2005).
 16. Giacomantone J., De Giusti A., ROC performance evaluation of RADSPM technique, XIV Congreso Argentino de Ciencias de la Computación (CACIC), Chilecito (2008).
 17. Chaovalitwongse W., Pardalos P. On the Time Series Support Vector Machine using Dynamic Time Warping Kernel for Brain Activity Classification, Cybernetics and Systems Analysis v.44 pp.125-138 (2008).
 18. Von Luxburg U. A Tutorial on Spectral Clustering. Statistics and Computing, 17(4), (2007).
 19. Shi J., Malik J. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 888-905, (2000).
 20. Lorenti L., Giacomantone J. Segmentación espectral de imágenes utilizando cámaras de tiempo de vuelo. XI Workshop Computación Gráfica, Imágenes y Visualización. pp. 430-439. Mar del Plata, Argentina, (2013).
 21. Yang, P.; Huang, B. An Outlier Detection Algorithm Based on Spectral Clustering. Pacific-Asia Workshop on Computational Intelligence and Industrial Application, vol.1, pp. 507-510, (2008).
 22. Han Y., Feng X., Baciú G. Variational and PCA based natural image segmentation. Pattern Recognition 46, pp. 1971-1984 (2013).
 23. Giacomantone J., Tarutina T. Diffuse Outlier Detection Technique for Functional Magnetic Resonance Imaging. Computer Science and Technology Series. XVI Argentine Congress of Computer Science Selected Papers. pp. 255-265 (2011).
 24. Penne et al. Robust real time 3-D respiratory motion detection using time of flight cameras. International Journal of Computer Assisted Radiology and Surgery v.3 pp.427-431. (2008).
 25. Kollors, A., Penne, J., Hornegger, J., Barker A. Gesture Recognition with a Time-of-Flight camera. International Journal of Intelligent Systems Technologies and Applications, v. pp.334-343, (2008).
 26. Li J., Wang Y. Visual tracing and learning using speeded up robust features. Pattern Recognition Letters 33, pp. 2094-2101 (2012).
 27. Prato M., Zanni L. A practical use of regularization for supervised learning with kernel methods. Pattern Recognition Letters 34, pp. 610-618 (2013).
 28. Antunez E., Marfil R., Bandera A. Combining boundary and región features inside the combinatorial pyramid for topology preserving perceptual image segmentation. Pattern Recognition Letters 33, pp. 2245-2253 (2012).
 29. Cristina Manresa-Yee, Ramón Mas, Gabriel Moyá, María J. Abásolo, Javier Giacomanto-

- ne. Interactive multi-sensory environment to control stereotypy behaviours. *Computer Science & Technology*, pp. 121-128, (2012).
30. Lorenti L., Violini L., Giacomantone J. Selección sub-óptima del espectro asociado a la matriz de afinidad. IV Workshop Procesamiento de Señales y Sistemas de Tiempo Real. pp. 1417-1425. Mar del Plata, Argentina, (2013).
 31. Giacomantone J., De Giusti A. Detección de áreas de interés bajo la hipótesis de relación espacial de voxels activados en fMRI. XX Congreso Argentino de Ciencias de la Computación. San Justo. Argentina. (2014).
 32. Lorenti L., Giacomantone J., De Giusti A. Segmentación de imágenes de tiempo de vuelo vía clustering espectral co-regularizado. XX Congreso Argentino de Ciencias de la Computación. San Justo. Argentina. (2014).

Herramientas de análisis de imágenes digitales para identificación y localización de objetos

María Lorena Bergamini, Jorge Kamlofsky

CAETI – Facultad de Tecnología Informática. Universidad Abierta Interamericana

Av. Montes de Oca 745. Ciudad de Buenos Aires

(+54 11) 43015323; 43015240 ; 43015248

María.Bergamini@uai.edu.ar; Jorge.Kamlofsky@uai.edu.ar

Resumen

En este trabajo se proponen e implementan herramientas de análisis de imágenes orientadas al desarrollo de algoritmos para reconocimiento, localización y clasificación de objetos en imágenes digitales.

Los objetos se pueden reconocer por su forma, y éstas se identifican definiendo sistemas de representación apropiados y métricas acordes, que deben ser de fácil y económica generación y manejo, a fin de utilizarse en tiempo real.

Se estudian representaciones compactas de formas que priorizan la reducción en la cantidad de datos a tratar, sin perder información acerca de la forma que está describiendo.

Se analizan alternativas de representación de configuraciones espaciales, y su incidencia en la proyección planar sobre una imagen digital bidimensional.

Palabras clave: visión artificial, invariantes geométricos, rectas digitales, curvatura digital, configuración espacial.

Contexto

El tema de investigación que abarca este proyecto está siendo estudiado por el grupo desde 2012. Se desarrolla en el Centro de Altos Estudios en Tecnología Informática (CAETI), dependiente de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana (UAI).

Los proyectos de investigación que se desarrollan en el CAETI se organizan en 5 líneas temáticas identificadas como prioritarias. El proyecto aquí presentado se enmarca en la línea prioritaria de Algoritmos y Software. Esta línea se orienta al diseño de software atendiendo el creciente aumento en la complejidad de las tareas a resolver, manteniendo calidad, tiempos y costos controlados. También se interrelaciona con proyectos de la línea prioritaria Automatización y Robótica. En esta línea se combinan mecánica, electrónica, física, informática, control, e inteligencia artificial para diseñar y manufacturar sistemas robóticos que puedan sustituir al ser humano en ciertas tareas.

Introducción

La identificación y localización automática de objetos en imágenes digitales mediante reconocimiento de

patrones, a partir de un sistema de visión artificial, permite realizar tareas como clasificación, vigilancia, manipulación, selección, navegación, etc.

El avance de los desarrollos tecnológicos permitió que esta capacidad se incorpore a dispositivos robóticos diseñados para fines específicos en una amplia variedad de aplicaciones que van desde la asistencia en la industria hasta el desempeño de tareas hogareñas.

Específicamente, entre estas tareas podemos mencionar control de calidad en la línea de fabricación, sistema de monitoreo de un predio para detección de intrusos, sistema automatizado de barreras en un estacionamiento privado, control de tránsito en determinada vía, conteo y seguimiento de objetos en movimiento, digitalización de información proveniente de soporte físico (papel, cartelería, patentes, etc), dactiloscopia, asistencia al diagnóstico médico por imágenes, etc.

La estimación automática de la posición de un objeto rígido moviéndose en el espacio tridimensional, a partir de imágenes bidimensionales del objeto, o a partir de datos proporcionados por sensores ubicados en el mismo, constituye un problema de fundamental importancia en la tecnología actual. Y junto con la tarea de determinar la posición, surge la necesidad de gobernar y controlar el movimiento automáticamente (Murray et al., 1994).

Para imágenes digitales que representan objetos tridimensionales, la tarea de reconocimiento es desafiante, ya que se cuenta con una proyección planar del mismo, perdiendo una dimensión. Esta representación parcial del objeto se puede enriquecer con proyecciones en distintas direcciones (multicámara), lo que aporta más información (sin llegar a reproducir completamente el objeto real). Más desafiante aún es la tarea cuando hay

problemas de iluminación, oclusión, ruido, etc.

El desarrollo de herramientas matemáticas para cumplir esta tarea ha sido foco de investigaciones de matemáticos e informáticos dedicados a visión computacional, videojuegos y robótica. Estas aplicaciones requieren formas eficientes de representar transformaciones de coordenadas en el espacio; sobre todo en operación y control online de cámaras o dispositivos robóticos.

Dado que las imágenes digitales, sobre todo en secuencias de tiempo real, involucran una alta cantidad de información digital, no es razonable aplicar métodos basados en análisis pixel a pixel, que requerirían tiempos computacionales incompatibles con el tiempo real. Es fundamental el diseño de algoritmos optimizados que puedan realizar tareas de reconocimiento de objetos manejando eficientemente la mínima información necesaria de la imagen.

El desarrollo de herramientas de soporte para la detección automática de objetos en tiempo real es un área que se ve favorecida por el crecimiento de la capacidad computacional, que permite alta velocidad de procesamiento y el aumento del poder de almacenamiento. En general, los algoritmos se estructuran en dos fases: aprendizaje y testeo. Se han desarrollado diversas estrategias para cada una de estas fases. Sin embargo, ningún algoritmo para reconocimiento de objetos es 100% preciso y óptimo (Prasad, 2012).

El reto es mejorar los procedimientos involucrados en visión artificial para lograr desempeños correctos en las distintas aplicaciones.

Una noción importante en visión artificial es similaridad de forma. Se han definido medidas de similaridad entre

formas para aplicar en algoritmos basados en aprendizaje por entrenamiento. Muchos investigadores han propuesto medidas de similaridad, o distancias entre formas, basados en una representación compacta de la misma [Latecki & Lakämper, 2000; Ling & Jacobs, 2007; Kamlofsky & Bergamini, 2013].

En una imagen digital, los puntos de máximo gradiente de tono constituyen el borde de una figura. El reconocimiento por análisis de bordes requiere eficientes procedimientos de extracción de características geométricas: curvatura, segmentos rectos y zonas de convexidad y concavidad. Estos procedimientos deben ser suficientemente insensibles al ruido presente en imágenes reales, a fin de obtener resultados confiables.

Una tarea central en geometría digital es el reconocimiento de rectas digitales (Ucska-Wehlou, 2009). Relacionado con esto, un concepto clave en el estudio geométrico de imágenes digitales es el de curvatura.

Varias generalizaciones de la definición de curvatura proveniente de geometría diferencial a la geometría digital han sido sugeridas. La definición de curvatura para curvas digitales ha dado lugar a diversos estimadores (Hermann & Klette 2007; Kerautret & Lachaud, 2008; Liu & Latecki, 2008). Varios de estos estimadores de curvatura se basan en medir segmentos digitales, lo cual es muy sensible a la presencia de ruido en la imagen (Nguyen & Debled-Renesson, 2007).

En aplicaciones en tiempo real (por ejemplo, reconocimiento de intrusos en una cámara de seguridad, control de ingreso y egreso vehicular en rutas) el uso de los recursos de procesamiento y almacenamiento debe ser eficiente.

En Kamlofsky y Bergamini (2013) se propone un método de aproximación de curvas digitales por polígonos, y se define

una representación compacta de la forma, a través del patrón de curvatura acumulada, lo que disminuye notablemente la cantidad de información a tratar al hacer la comparación de formas para su reconocimiento. Además, se propone una medida de similaridad de formas, basada en la curvatura acumulada.

Hasta aquí, en el grupo de investigación, se han estudiado objetos a través de su representación bidimensional en el plano. Nos enfrentamos ahora a la tarea de incorporar la tercera dimensión, para la identificación y localización, considerando volúmenes, ubicación en el espacio, yuxtaposición, oclusión, etc.

Es necesario operar con un óptimo sistema de representación de configuraciones espaciales. Antes esta cuestión surgen diversas parametrizaciones posibles, que se fundamentan principalmente en matrices de rotación, ángulos de Euler, cuaterniones, y otras derivadas de estos elementos.

Existen varias alternativas para representar posiciones o movimiento 3D, y las distintas estrategias implican ventajas y desventajas de implementación que deben ser tenidas en cuenta (Kamlofsky & Bergamini, 2015a, 2015b).

Líneas de Investigación, Desarrollo e Innovación

El objetivo de este proyecto es generar conocimiento teórico y práctico en el área de análisis de imágenes. En un proyecto anterior se centró el estudio en objetos digitales 2D, y en esta continuación se pretende estudiar los invariantes geométricos en la representación bidimensional de objetos tridimensionales en imágenes digitales.

En el aspecto teórico se estudiarán cuestiones de geometría digital, geometría proyectiva, localización espacial, definición, extracción y cálculo de invariantes geométricos y descriptores de objetos en una imagen digital; distintas parametrizaciones de configuraciones espaciales. Desde el punto de vista práctico se pretende implementar dicho conocimiento teórico en algoritmos que permitan llevar a cabo reconocimiento y localización automáticos y eficientes de objetos en una escena tridimensional, para ser aplicados a sistemas de visión artificial.

Se estudian descriptores invariantes frente a transformaciones afines y proyectivas, y se definen medidas de similaridad relacionados con cada uno, o con ciertos conjuntos pertinentes de descriptores.

Los resultados teóricos obtenidos fundamentan un algoritmo de búsqueda, reconocimiento, localización y clasificación de objetos, que será diseñado e implementado para tal fin.

Resultados y Objetivos

Los objetivos específicos son

a) Estudiar en forma teórica la definición de características geométricas en objetos digitales 3D.

b) Estudiar las distintas herramientas matemáticas para el cálculo de configuraciones de objetos en una escena.

b) Estudiar la preservación de propiedades de un objeto mediante ciertas transformaciones geométricas (afines y proyectivas).

c) Proponer métodos de representación simplificada de objetos y medidas de similaridad, compatibles con el procesamiento en tiempo real.

d) Estudiar, desarrollar y comparar técnicas de representación de posiciones y movimientos en 3D

e) Aplicar las técnicas y métodos propuestos para la determinación de ubicación espacial a partir de análisis de imágenes.

e) Diseñar algoritmos de reconocimiento y clasificación de objetos en imágenes digitales.

Hasta el momento, en el grupo de investigación se ha desarrollado un método de representación de formas digitales, basado en la aproximación de la forma del objeto por un polígono y obtener luego una representación de la evolución de la curvatura a lo largo de todo el contorno. Dicha representación simplificada permite realizar comparaciones entre curvas digitales, de modo de poder obtener una medida de similitud entre formas.

Un concepto central en la aproximación de formas por polígonos es el de segmento recto digital (SRD). Se han estudiado caracterizaciones algebraicas y sintácticas de SRD y de barras digitales (generalización del concepto de SRD). Este tema está siendo estudiado por un tesista de Licenciatura en Matemática.

A partir del polígono aproximante, se genera el patrón de giro, que involucra la idea de la curvatura acumulada. Este patrón contiene toda la información necesaria para extraer características geométricas de la forma analizada, ya que resulta ser invariante frente a traslaciones, simetrías y escalado uniforme.

La invariancia frente a rotaciones se logra identificando puntos de curvatura extrema (sharp points). Esto permite dividir el contorno en partes significativas.

Otro enfoque para la poligonalización de una curva digital es mediante la cobertura tangente. Esto implica determinar arcos de curva que son tangentes de todos sus puntos. Estas ideas

constituyen el tema de estudio de una tesis de Licenciatura en Matemática.

Formación de Recursos Humanos

Este proyecto es la continuación de uno desarrollado en el período 2014-2016. En ese sentido, contamos con experiencia en el tema objeto de estudio. Los integrantes del proyecto tienen una sólida formación en matemática, necesaria para enfocar los objetivos propuestos. Hay dos alumnos de Licenciatura en Matemática desarrollando su tesis de grado en esta área, y un alumno de posgrado finalizando su tesis de Maestría en Tecnología Informática, y desarrollando su tesis de Doctorado en Ingeniería.

Referencias

- Hermann S., Klette R. “A comparative study on 2D curvature estimators”. *Int. Conf. on Comp.: Theory and Applications*, pp. 584-589, 2007.
- Kamlofsky J., Bergamini M.L. “Cuaterniones en visión 3D”. *Revista MACI (Matemática Aplicada, Computacional e Industrial)*, Vol 5, pag. 517-520, 2015a.
- Kamlofsky J., Bergamini M.L. “Parametrizaciones de Movimientos Rígidos en 3D para Visión Artificial”. *CONAISI 2015*, UTN, FRBA, 2015b.
- Kamlofsky, J. Bergamini M. L. “Patrón de Evolución de Angulo de Giro para el Reconocimiento de Objetos en Imágenes Digitales”, *JAIIO*, 2013.
- Kerautret B., Lachaud J.-O. “Curvature Estimation along Noisy Digital Contours by Approximate Global Optimization”. *Patt. Recogn.* 42(10), pp. 2265 – 2278, 2008.
- Latecki L. J., Lakämper R. “Shape Similarity Measure Based on Correspondence of Visual Parts”, *IEEE Trans. Pattern Analysis and Machine Intelligence* 22(10), pp. 1185-1190, 2000.
- Ling H, Jacobs D. “Shape Classification Using the Inner-Distance”, *IEEE Trans. Pattern Analysis and Machine Intelligence* 29(2), pp. 286-299, 2007.
- Liu H. R., Latecki L. J., Liu W. “A Unified Curvature Definition for Regular, Polygonal, and Digital Planar Curves”. *Int. J. Comput. Vis.* 80, pp. 104–124, 2008.
- Murray R., Li Z., Sastry S. “A Mathematical Introduction to Robotic Manipulation”. *CRC Press*, Inc. Boca Raton, FL, USA, 1994.
- Nguyen T.P., Debled-Rennesson I. “Curvature Estimation in Noisy Curves”. In: Kropatsch W., Kampel M., Hanbury A. (eds.) *Computer Analysis of Images and Patterns, LNCS*, 4673, pp. 474-481, Springer, Heidelberg, 2007.
- Prasad D. “Survey of The Problem of Object Detection In Real Images”. *Int. J. of Images Processing* 6 (6), pp. 441-466, 2012.
- Ucska-Wehlou H. “Digital lines, Sturmian words and continued fraction”. *Uppsala Dissertations in Mathematics* 65, 2009.

La Realidad Virtual en los Comportamientos Sociales

N. Jofré, G. Rodríguez, Y. Alvarado, J. Fernández, R. Guerrero

LCG - Universidad Nacional de San Luis

Ejército de los Andes 950

Tel: 0266 4420823, San Luis, Argentina

{npasinetti, gbrodriguez, ymalvarado, jmfer, rag}@unsl.edu.ar

Resumen

Las nuevas tecnologías de la información permiten la generación de entornos de interacción que proveen nuevos mecanismos de comunicación e intercambio de información con una computadora; dentro de esta revolución tecnológica se encuentra la Realidad Virtual (RV).

La RV es una tecnología poderosa cuyas aplicaciones en la vida cotidiana son infinitas, es por ello que aborda tópicos cada vez más complejos entre los cuales se encuentran los comportamientos de los individuos en sociedad. La realidad virtual ha logrado tener un gran impacto psicológico gracias a que puede engañar al cerebro para que experimente sensaciones que no son reales, pero que lo parecen, permitiendo recrear situaciones con el objetivo de analizar, tratar e influir los comportamientos sociales. En este sentido, la RV permite incluso alcanzar un grado de exposición superior al que sería posible en vivo, dado que la manipulación de un entorno virtual facilita enfatizar aquellas situaciones a investigar; es por ello que la sensación de presencia es un aspecto clave.

Esta propuesta de trabajo pretende analizar los comportamientos sociales que podrían llegar a ser influenciados mediante el uso de realidad virtual, con la intención de incorporarla como herramienta alternativa para

afrontar problemáticas relacionadas a comportamientos sociales.

Palabras Claves: Realidad Virtual, Computación Gráfica, Interfaces Humano-Computadoras, Vida Artificial, Personajes Virtuales.

Contexto

La propuesta de trabajo se lleva a cabo dentro de la línea de Investigación “Procesamiento de Información Multimedia” del proyecto “Tecnologías Avanzadas Aplicadas al Procesamiento de Datos Masivos”. Este proyecto es desarrollado en el ámbito del Laboratorio de Computación Gráfica de la Universidad Nacional de San Luis.

1. Introducción

Hace algún tiempo los científicos predijeron la inminente transición en las formas en que los humanos experimentarían la comunicación de información. La realidad virtual ha permitido la generación de ambientes de interacción que facilitan nuevos contextos de intercambio y comunicación de información. Actualmente, las aplicaciones de RV consisten en la inmersión del usuario en un ambiente generado por computadora, dando lugar a una idea natural para mejorar la impresión de vivir en una realidad simulada [1, 2].

Es conocido que el uso de técnicas de RV no altera el cuerpo físico pero sí afecta a las

percepciones e interacciones del ser humano, ya que éste se desenvuelve eficientemente en un entorno virtual en tiempo real haciendo uso de sus sentidos y habilidades naturales. En sus orígenes, la RV fue concebida como una herramienta netamente tecnológica para entretenimiento. Cabe recordar al primer sistema de RV denominado *Sensorama*, el cual consistía en un demo de un paseo en bicicleta por Brookling [1]. Tiempo después el objetivo se orientó a desarrollar simulaciones de escenarios para entrenamiento (de guerra o civiles) [3]. Posteriormente, la amplia diversidad de usos de la RV permitió su extensión a otros campos como la salud y la educación [4]. Dentro de la salud, las aplicaciones actuales abarcan desde animaciones 3D para estudiar y tratar pacientes que padecen trastornos por estrés post-traumático, fobias y adicciones, hasta simulaciones de operaciones complejas para práctica de residentes y médicos experimentados [5, 6]. Dentro de la educación, los entornos de enseñanza-aprendizaje se realizan en aulas virtuales, donde a través de tecnología de red, es posible recrear situaciones de aprendizajes individuales y/o grupales, tales como competencias para alumnos y prácticas pedagógicas para docentes [7, 8].

Como se sabe, el ser humano es un individuo que constantemente se encuentra en un entorno relacional. Hoy por hoy, el mundo atraviesa una etapa conocida como *cibercultura*, la cual consiste de comunidades virtuales que contribuyen a la socialización de las personas propiciando un flujo de información ágil entre ellas. Dicha cultura está altamente influenciada por el uso masivo de redes sociales y mundos virtuales [9]. Ambos, son productos innovadores que aprovechan las sinergias colectivas, es decir, surgieron debido a las innumerables situaciones de la vida y las necesidades propias del ser humano. Es por ello que el uso de la tecnología y en particular de la RV implica el análisis de otros factores tales como los sociales y los psicológicos [10].

Como consecuencia, últimamente se ha potenciado el proceso de consideración de uso de la RV para el estudio y mejora de las habi-

lidades sociales, la cognición, y el desempeño social. Aprovechando la eficacia de la RV en su capacidad para cerrar las brechas tanto perceptivas como temporales, es que muchas empresas de juegos de computadora han optado por basar sus juegos en conductas sociales tales como los de tipo *Second Life* [11]. Un ejemplo es el juego *Los Sims*, el cual es un videojuego de simulación social y de estrategia en el que cada ser virtual tiene personalidad propia y se controla individualmente de forma directa [12]. De la misma forma han surgido otros juegos donde el ambiente y la banda sonora influye sobre la psicología del usuario. Dentro de estos juegos se encuentran los mundialmente conocidos *Silent Hill*, *Tomb Raider*, *Dead Space*, *Eternal Darkness*, *The Evil Within* [13, 14, 15, 16].

El desarrollo de estos tipos de juegos por parte de la industria del entretenimiento y su aplicación en el ámbito científico, ha permitido que la ciencia pueda analizar la simulación de diferentes situaciones del mundo real y el correspondiente comportamiento de las personas en su experiencia con el entorno simulado (terapias cognitivo-conductuales) [17, 18]. Del análisis se desprenden el estudio de la *comunicación interpersonal*, la *asertividad*, el *reconocimiento de emociones*, la *ansiedad social*, la *fobia*, entre otros. Dependiendo de la situación simulada serán las reacciones experimentadas por parte del usuario: 1) si el usuario se encuentra en ambientes cotidianos como una plaza, un banco, la calle, etc., se espera que el ambiente exponga el comportamiento normal de la persona (grado de civismo) [19]; 2) si el usuario se encuentra en un entorno hostil (estilo *Tomb Raider*), es posible que deba explorar y superar sus propios miedos (soledad, altura, etc.) [20], algo totalmente diferente a si el usuario es expuesto a situaciones inesperadas (aleatorias), donde éste debería no sólo superar sus miedos ya conocidos sino también los miedos desconocidos; con lo cual su comportamiento sería más parecido al de un soldado en un campo de batalla, donde tiene que hacer uso de lo aprendido durante su entrenamiento y de su capacidad innata para afrontar situaciones desconocidas.

Las tecnologías de RV son una alternati-

va interesante para el análisis y evaluación de los diferentes comportamientos del ser humano ante diversas situaciones. Cada vez más, la RV está demostrando ser una herramienta fundamental en la implementación y práctica de cambios sociales mundiales tales como la aceptación de la diversidad, el respeto por el medio ambiente, la solidaridad, etc.

2. Líneas de Investigación y Desarrollo

En función de lo anteriormente expresado, resulta que la realidad virtual se convierte en una herramienta poderosa que posibilita asistir y mejorar las conductas o comportamientos sociales en diversas áreas. Se considera de principal interés a aquellas áreas donde se sabe o se espera que la RV puede aportar grandes beneficios como lo son: el análisis del comportamiento cívico en ambientes cotidianos, el comportamiento de superación en ambientes de riesgo pero a la vez controlados, y el comportamiento de entrenamiento en ambientes de riesgo extremo en situaciones caóticas.

- *Comportamiento cívico.* Sabiendo que el civismo es, por encima de todo, la cultura de la convivencia pacífica y solidaria, del compromiso con su ciudad y sus habitantes, se pueden simular escenarios virtuales cotidianos de la vida en sociedad y recrear situaciones para educar a las personas a lograr la mencionada convivencia. Es decir, vivir en comunidad respetando normas de convivencia pacífica, aceptando las reglas del juego de la democracia y los derechos fundamentales o los valores constitucionales. Al ofrecer distintas actividades en forma creciente, en cuanto a su complejidad en interacción con dispositivos de RV y valores morales que se pretenden enseñar o reforzar, se espera brindar a las personas un conjunto de cualidades que les permita vivir en sociedad. Estas actividades pueden ser: recomendar cruzar a un peatón por la esquina y no a mitad de cuadra, recoger

residuos del piso y llevarlos hasta un cesto de basura o, desde el punto de vista de un conductor, ceder el paso a los peatones, entre otros. Dentro de un ambiente virtual, la tecnología de RV permite obtener información sobre la localización y situación de las personas. De esa manera, el usuario dispone de las herramientas que le permitirán prestar atención hacia su entorno [21, 22, 23].

- *Comportamiento de superación.* La RV puede ayudar a las personas a superar o enfrentar problemas comunes que requieren la atención terapéutica. Los tratamientos psicológicos para los trastornos como las fobias implican algún tipo de exposición a los estímulos temidos, ya que de las teorías explicativas conductuales se desprende que dicha exposición en ausencia de las consecuencias temidas tendrá como resultado la extinción de las reacciones fóbicas. Dado que la RV nos permite introducirnos en ambientes virtuales, podemos recrear situaciones que simulen la exposición de las personas a sus fobias con la total seguridad de no estar realmente expuestas. Comenzando con un nivel de exposición bajo y observando las respuestas del paciente a través de dispositivos especializados, se puede elevar gradualmente el nivel hasta lograr una exposición casi real a la fobia. De esta manera, la RV se convierte en una herramienta ideal para ayudar a las personas con este o incluso otros problemas de la vida cotidiana [24] [25] [26].
- *Comportamiento de entrenamiento.* Todos los ciudadanos hombres, mujeres, niños y niñas tienen la obligación moral de prepararse en la educación de la defensa civil para poder ayudarse unos a otros en caso de emergencia; así como también acudir en apoyo de todas las personas, instituciones y sociedad en general. La RV puede organizar, instruir, preparar y capacitar a los habitantes a fin de que puedan afrontar en forma adecuada

da toda situación de emergencia producida por conflictos de tipo internacional, conmoción interna y desastre imprevisible provocados por la naturaleza o por el hombre. Mediante distintos dispositivos de RV se puede reconocer la situación de una persona (tanto su capacidad física como también calcular la probabilidad de poder realizar acciones exitosamente), y brindar una lista de sugerencias en tiempo real (a través de pequeñas frases o palabras claves transmitidas por un sistema de sonido) de las acciones posibles más prometedoras, con el fin de proteger al prójimo y protegerse a sí mismo [27] [28].

3. Resultados obtenidos / esperados

El grupo de trabajo, además de pertenecer a un proyecto de investigación de la UNSL, ha desarrollado tareas dentro del marco de un Proyecto ALFA III de la Comunidad Europea, denominado GAVIOTA (Grupos Académicos para la Visualización Orientada por Tecnologías Apropriadas), en el que participó la UNSL en conjunto con otras universidades de América Latina y Europa. Asimismo, se ha obtenido subsidio por parte de la Secretaría de Políticas Universitarias para la concreción de las actividades ya iniciadas.

Como consecuencia, se ha desarrollado un personaje virtual del tipo pregunta-respuesta, el cual ha sido contextualizado dentro de un sistema de RV inmersivo, y un juego bajo la filosofía de *crowdsourcing* para la influencia de cambio de comportamiento cívico.

Actualmente las acciones se encuentran focalizadas en la incorporación de nuevas técnicas cognitivo-conductuales para el análisis de los diversos comportamientos sociales.

4. Formación de Recursos Humanos

Los trabajos realizados han permitido la concreción de trabajos de fin de carrera de la Licenciatura en Ciencias de la Computación, así como también la definición de un trabajo de tesis de Maestría en Ciencias de la Computación, actualmente en ejecución. Adicionalmente se ha obtenido una beca de iniciación a la investigación otorgada por la Secretaría de Ciencia y Técnica de la UNSL.

Referencias

- [1] W. R. Sherman and A. B. Craig. *Understanding VR: Interface, Application, and Design*. Morgan Kaufmann, SF, CA, USA, 2002.
- [2] J. Fox, D. Arena, and J. N. Bailenson. VR: A survival guide for the social scientist. *Journal of Media Psychology: Theories, Methods, and Applications*, 21(3):95–113, 2009.
- [3] A. Craig, W. R. Sherman, and J. D. Will. *Developing VR Applications: Foundations of Effective Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2009.
- [4] R.J. Lamson. VR immersion therapy for treating psychological, psychiatric, medical, educational and self-help problems, July 30 2002. US Patent 6,425,764.
- [5] T. D. Parsons and A. Rizzo. Virtual Human Patients for Training of Clinical Interview and Communication Skills. In *International Conf on Disability, VR and Associated Technology*, Maia, Portugal, September 2008.
- [6] Davide Corbetta, Federico Imeri, and Roberto Gatti. Rehabilitation that incorporates vr is more effective than standard rehabilitation for improving walking speed, balance and mobility after stroke: a systematic review. *Journal of Physiotherapy*, 61(3):117 – 124, 2015.
- [7] I. Cano, A. Gallardo, and J. Salinas. Un entorno virtual de apoyo a la investigación e intercambio de materiales y experiencias de formación. In *VII Congreso Internacional de NTICs para la Educación*, 2004.

- [8] D.H. Choi. *Emerging Tools and Applications of Virtual Reality in Education*. Advances in Educational Technologies and Instructional Design. IGI Global, 2016.
- [9] Ralph Schroeder. Cyberculture, cyborg post-modernism and the sociology of virtual reality technologies: Surfing the soul in the information age. *Futures*, 26(5):519 – 528, 1994.
- [10] Ralph Schroeder. *Possible Worlds: The Social Dynamic of Virtual Reality Technology*. Westview Press, Inc., Boulder, CO, USA, 1996.
- [11] Tom Boellstorff. *Coming of Age in Second Life: An Anthropologist Explores the Virtually Human*. Princeton University Press, Princeton, NJ, USA, 2008.
- [12] Gonzalo Frasca. The sims: Grandmothers are cooler than trolls. *Game Studies*, 1(1), 2001.
- [13] Gozo Kitao. Silent hill. In *ACM SIGGRAPH 99 Electronic Art and Animation Catalog, Los Angeles, CA, USA, August 8-13, 1999*, page 188, 1999.
- [14] R. Sifa, A. Drachen, C. Bauckhage, C. Thureau, and A. Canossa. Behavior evolution in tomb raider underworld. In *2013 IEEE Conference on Computational Intelligence in Games (CIG), Niagara Falls, ON, Canada, August 11-13, 2013*, pages 1–8, 2013.
- [15] *Dead Space*. <http://www.ea.com/dead-space-3>, 2013.
- [16] *The Evil Within*. <http://theevilwithin.com>, 2014.
- [17] Bruno Herbelin, Françoise Riquier, Frederic Vexo, and Daniel Thalmann. Virtual reality in cognitive behavioral therapy: a study on social anxiety disorder. In *In 8th International Conference on Virtual Systems and Multimedia (VSMM)*, 2002.
- [18] B Herbelin, P Benzaki, F Riquier, O Renault, and D Thalmann. Using physiological measures for emotional assessment: a computer-aided tool for Cognitive and Behavioural Therapy. In *ICDVRAT 2004*, pages 307–314, 2004.
- [19] A. van Hamel. *From Consumer to Citizen: Digital Media and Youth Civic Engagement*. 2014.
- [20] Willem-Paul Brinkman, Charles van der Mast, and Daniel de Vliegher. Virtual reality exposure therapy for social phobia: A pilot study in evoking fear in a virtual world. In *Proceedings of HCI2008 workshop - HCI for technology enhanced learning*, page 85–88, 2008.
- [21] Victòria Camps. El sentido del civismo. *Los monográficos de Barcelona Metròpolis Mediterrànea B. MM*, (6):15–21, 2005.
- [22] N. Jofré, G. Rodríguez, Y. Alvarado, J. Fernández, and R. Guerrero. A serious game about recycling rules. In Universidad Nacional del Noroeste de la Provincia de Buenos Aires, editor, *XXI Congreso Argentino de Ciencias de la Computación*, pages 262–271, Octubre 2015.
- [23] N. Jofré, G. Rodríguez, Y. Alvarado, J. Fernández, and R. Guerrero. Natural interface for recycling activity. In *6TH International Symposium on Innovation and Technology*, Agosto 2015.
- [24] José Gutiérrez Maldonado. Aplicaciones de la realidad virtual en psicología clínica. *Aula médica psiquiatría*, 4(2):92–126, 2002.
- [25] Cristina Botella Arbona, Azucena García-Palacios, and Rosa María Baños. Realidad virtual y tratamientos psicológicos. *Cuadernos de medicina psicosomática y psiquiatría de enlace*, (82):17–31, 2007.
- [26] C. Botella, R. Baños, C. Perpiñá, S. Quero, H. Villa, A. García Palacios, and S. Febregat. El tratamiento de la claustrofobia por medio de realidad virtual. *Análisis y Modificación de conducta*, 28(117):109–127, 2002.
- [27] Mara Malaver. Programa de capacitación a distancia en primeros auxilios y prevención ante efectos destructivos de eventos naturales.
- [28] Alejandro Fernández Arroyo et al. Implementación de situaciones de riesgo en un simulador de accidentes de tráfico mediante motor de videojuegos unity. 2015.

Realidad Aumentada, Realidad Virtual, Interfaces Avanzadas, Juegos Educativos

María José Abásolo^{1,2}, Alejandro Mitaritonna³, María José Bouciguez¹,
Natalia Encina¹, Mario Vicenzi¹, Armando De Giusti¹, Marcelo Naiouf¹,
Javier Giacomantone¹, Cristina Manresa-Yee⁴

¹Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata (UNLP)
{mjabasolo, degiusti, mnaiouf, jog}@lidi.info.unlp.edu.ar

²Comisión de Investigaciones Científicas de la Provincia de Buenos Aires
(CICPBA)

³ Instituto de Investigaciones Científicas y Técnicas para la Defensa (CITEDEF)
amitaritonna@citedef.gob.ar

⁴ Departamento de Cs. Matemáticas e Informática – Universidad de las Islas
Baleares
cristina.manresa@uib.es

Resumen

La línea de investigación y desarrollo presentada consiste en estudiar, desarrollar y evaluar aplicaciones de realidad virtual, realidad aumentada e interfaces avanzadas. Uno de los principales objetivos es la formación de recursos humanos y fortalecimiento de la investigación mediante el trabajo intergrupar entre diferentes instituciones nacionales y extranjeras. El proyecto concluido Formación de Recursos Humanos e Investigación en Visión e Informática Gráfica (FRIVIG) de la AECI permitió la compra de equipamiento, el intercambio de investigadores, la formación de recursos humanos, y la creación de una carrera de postgrado en la temática.

Palabras Clave: Realidad Virtual, Realidad Aumentada, Visión por Computador, Informática Gráfica, Interfaces Avanzadas, Juegos Educativos

Contexto

La línea de investigación Realidad Aumentada y Realidad Virtual forma parte del proyecto 11/F017 "Cómputo Paralelo de Altas Prestaciones. Fundamentos y Evaluación de rendimiento en HPC. Aplicaciones a Sistemas Inteligentes, Simulación y Tratamiento de Imágenes", acreditado por la Universidad Nacional de La Plata (UNLP) en el marco del Programa de Incentivos. Dentro de esta línea se está desarrollando una tesis doctoral en el Instituto de Investigaciones Científicas y Técnicas para la Defensa (CITEDEF) en el marco del Programa de Investigación y Desarrollo para la Defensa (PIDDEF) (aprobado con el Nro. 22/12), elaborado por la Subsecretaría de Investigación Científica y Desarrollo Tecnológico del Ministerio de Defensa.

Introducción

La línea de investigación y desarrollo presentada involucra las áreas de visión por computador e informática gráfica.

Los temas de investigación estudiados se agrupan en cuatro ejes temáticos: realidad virtual, realidad aumentada, interfaces avanzadas y juegos educativos. En [Man12a] puede encontrarse una introducción a esta área presentada por el grupo de investigación en el marco parte de una escuela de computación del congreso nacional CACIC.

Realidad Aumentada

La Realidad Aumentada (RA) se refiere a aplicaciones interactivas en tiempo real donde se visualiza la realidad con elementos sintéticos agregados (objetos 3D, sonidos, texto, etc.) de forma coherente con el punto de vista del usuario. Recientemente el uso de RA se ha combinado con la geolocalización del participante, es decir, la localización de su ubicación en el globo terráqueo, en base a dispositivos físicos como GPS, brújulas, acelerómetros y giroscopios incorporados a celulares de última generación. De esta manera, la combinación de RA y geolocalización, dio lugar a las herramientas de software de RA denominadas browsers de RA. Definidas por Lester Madden [Mad11] como aplicaciones desarrolladas para dispositivos móviles con dispositivos de geolocalización incorporados, que proveen contenidos relevantes del entorno -en forma de objetos 3D, imágenes, indicadores de puntos de interés (POIs) entre otros- dependiendo de la ubicación del usuario en un cierto lugar en el mapa terrestre .

El desafío de aprender a incorporar la RA y la geolocalización en los contextos educativos, implica la voluntad de centrar el aprendizaje en la participación activa del estudiante, en sus intereses, en situaciones relevantes y directamente relacionadas con su vida real, lo cual supone un cambio en los planteamientos pedagógicos que exigen el diseño de nuevas propuestas

metodológicas y el uso de recursos didácticos capaces de facilitar los nuevos procesos. Pero para ello, necesitamos docentes que conozcan y aprendan a utilizar estas herramientas, para luego mediar el contenido de su disciplina desde una mirada pedagógica – didáctica, pudiendo vislumbrar las posibilidades educativas de la RA y la geolocalización para enriquecer los procesos de enseñanza – aprendizaje; brindando escenarios para explorar, interactuar y relacionarse con su entorno, generando recursos y conocimientos de manera creativa y lúdica. En el ámbito educativo, se ha trabajado sobre la evaluación de los browsers de RA más relevantes como por ejemplo Layar, Wikitude y espiRA.

En el ámbito militar, durante el año 2015 se ha trabajado sobre la implementación del framework de RA denominado RAIOM (Realidad Aumentada para la Identificación de Objetivos Militares) desarrollado en el CITEDEF (Instituto de Investigaciones Científicas y Técnicas para la Defensa)[Mit13][Mit15]. Dicho framework fue utilizado para desarrollar una aplicación en Android que integrando librerías OpenCV y se ejecuta sobre un Smartphone Nexus. Entre las funcionalidades del framework de RA se enumeran:

- Visualización de objetivos tácticos en un radar en 360 grados (figura 1)
- Posicionamiento y orientación espacial de los objetos y del operador (figura 1)
- Ayudas visuales mediante iconografía/simbología adaptable (figura 1)
- Uso de menús interactivos mediante reconocimiento gestual (figura 2 y figura 3)
- Uso de cartografía en 3D mediante uso de Realidad Aumentada (figura 4)



Figura 1

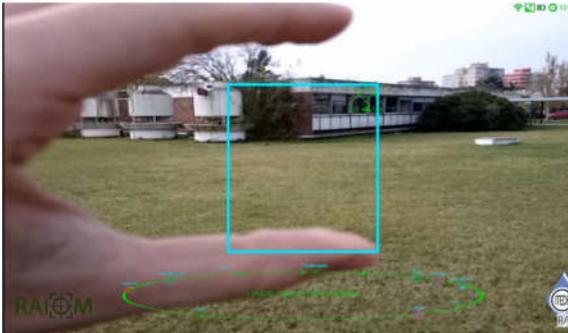


Figura 2

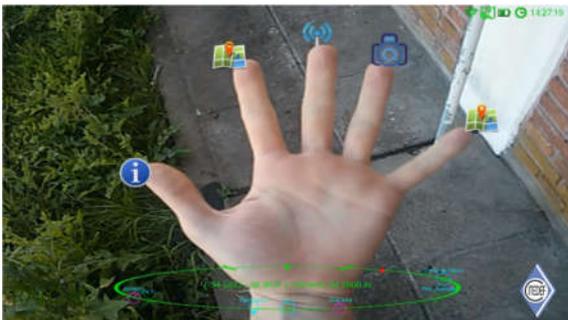


Figura 3

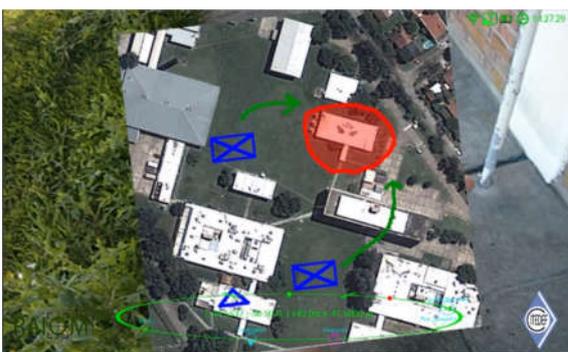


Figura 4

Entre las tareas actuales es la implementación de otra aplicación que se ejecute directamente sobre los microprocesadores ODROID-XU y utilizando como medio de visualización las gafas de RA.

Realidad Virtual

La Realidad Virtual (RV) es un término que se aplica a un conjunto de experiencias sensoriales sintéticas, es decir generadas por computador, comunicadas a un operador o participante. La mayoría de las aplicaciones de realidad virtual son experiencias visuales donde el participante se ve inmerso e interactúa en un ambiente o escena virtual, con diferentes grados de inmersión.

Algunas aplicaciones de realidad virtual requieren la visualización de terrenos. La principal dificultad a la que se enfrentan las técnicas de visualización de terrenos es obtener una visualización eficiente en tiempo real para lograr terrenos que luzcan realistas. La visualización en tiempo real impone restricciones estrictas en la eficiencia de los sistemas de visualización, por lo que es necesario a su vez negociar la calidad de la visualización con un uso limitado de recursos. Es por esta razón que los algoritmos que han sido propuestos utilizan el concepto de multiresolución para generar representaciones simplificadas, por ejemplo, realizando una representación de menor nivel de detalle o LOD (Level Of Detail) a medida que la superficie del terreno está más lejos del punto de observación. Hoy en día, es posible implementar algoritmos que serán ejecutados directamente por las unidades de procesamiento gráfico (GPUs). Los últimos avances en las capacidades de procesamiento gráfico han introducido la posibilidad de realizar teselados de geometría directamente por hardware (GPU), y con una mínima intervención por parte de la CPU. Borelli y Abásolo [Bor15] presentan un nuevo algoritmo de visualización de terrenos multiresolución que emplea los últimos avances de la GPU. En su diseño se aplica un criterio de selección de nivel de detalle que considera la percepción que tendrá el usuario de la rugosidad

particular de cada zona del terreno. Además, se incorporan las capacidades de teselado por hardware de la GPU correspondientes a Shader Model 5, y se implementa un mecanismo de geomorphing creando transiciones suavizadas entre distintos niveles de detalle. La técnica realiza una representación multiresolución eficiente y escalable. Los resultados obtenidos posibilitan su utilización en aplicaciones interactivas que requieren de la visualización de terrenos en tiempo real para aplicaciones 3D.

La aplicación realizada se probó utilizando una pantalla panorámica conformada por 3 TVs de alta resolución Full HD, una a continuación de la otra, formando un ángulo de 120° entre pantallas contiguas. Se utilizó la placa de video GeForce GTX 690 capaz de administrar 3 pantallas de manera transparente al usuario.

Interfaces Avanzadas para discapacitados

La interacción persona-ordenador (IPO) es la disciplina relacionada con el diseño, la evaluación y la implementación de sistemas informáticos interactivos para que los utilicen los seres humanos [Hew92]. El objetivo principal de la disciplina es conseguir una interacción eficiente, efectiva y satisfactoria, comprendiendo los factores que determinan cómo son las personas y cómo actúan y desarrollando herramientas y técnicas a aplicar en el Diseño Centrado en el Usuario (DCU) [ISO10].

La interacción ha evolucionado desde interfaces difíciles de aprender como las líneas de comandos hasta interfaces más intuitivas como las interfaces naturales (p.e. pantallas multitáctiles o interfaces gestuales). Tanto la investigación como el mundo empresarial están explorando nuevos canales de interacción, y en esta línea se centran las interfaces avanzadas, que

incluyen, entre otras, interfaces hápticas, interfaces gestuales o interfaces cerebro-ordenador.

Uno de los sectores de la población que pueden beneficiarse de estas nuevas interfaces son las personas con discapacidad, especialmente a través de diseñar sistemas interactivos que tengan en cuenta la accesibilidad. En este dominio de investigación, Manresa et al. [Man13] presentan un sistema educativo interactivo para personas con discapacidades múltiples y severas que necesitan ayuda en todas las actividades diarias pero que muestran interés cuando aparece un estímulo significativo en el entorno. Los objetivos del sistema son mejorar su interacción con el entorno, ofrecerles control sobre el mismo, motivar los movimientos de las extremidades superiores o reducir sus movimientos de auto-estimulación [Man12b]. El sistema utiliza visión por computador para detectar una banda de color sobre la mano del usuario y transforma esa posición en una posición sobre la pantalla. En la pantalla hay diferentes regiones que pueden ofrecer estímulos diversos (audio, imagen, vídeo) y cuando se detecta al usuario posicionado en una región activa, se inicia el estímulo que parará cuando el usuario salga de esa región.

Al evaluar el sistema anterior, se observa que el terapeuta tiene que motivar oralmente a los usuarios y a veces ayudarlos físicamente (tocando el brazo a levantar, o ayudando al usuario a realizar el movimiento). Para intentar evitar la ayuda física, Manresa et al. [Man15] presentan una interfaz vibrotáctil que pretende motivar la flexión del brazo a través de explotar la ilusión perceptual del “conejo cutáneo”, donde una sucesión rápida de vibraciones cerca del codo y yendo hacia el hombro, crea la sensación de golpecitos hacia arriba incluso en zonas donde no ha habido estímulo físico, por

lo que se puede informar de direccionalidad y llamar la atención del usuario para que realice el movimiento.

Bojanovic et al. [Boj15] trabajan con el mismo perfil de usuario y con los mismos objetivos. La interacción se basa en la información que proviene de un sensor de distancia cuando el usuario acerca o aleja la mano, y la transmite a seis aplicaciones de acción/reacción que ofrecen retroalimentación significativa y multimodal (imágenes, audio, vídeo).

Juegos Educativos

Balasubramanian y Wilson [Bal06] definen juego un ambiente interactivo atractivo que cautiva al jugador ofreciéndole desafíos que requieren niveles crecientes de maestría. El término juego educativo hace referencia a aquellos juegos desarrollados con finalidad educativa. Características como la diversión, la inmersión y la interactividad, hacen a los videojuegos una herramienta digital atractiva para el ámbito educativo.

El aprendizaje basado en juego digital (DGBL, Digital Game Based Learning) es un método de enseñanza que incorpora contenidos educativos o principios de aprendizaje en los videojuegos, con el objetivo de involucrar a los estudiantes [Squ08]. Los videojuegos educativos son aplicaciones de software en las que el estudiante aprende haciendo, interactuando con el mundo, observa los resultados de sus acciones, y prueba sus hipótesis.

Los videojuegos, generalmente, tienen lugar en una especie de mundo virtual estructurado por normas específicas, con mecanismos de retroalimentación y herramientas de apoyo. Las simulaciones comparten con los juegos el uso de un mundo virtual, estructurado por reglas y restricciones, y la focalización en un objetivo concreto; aunque emplean un conjunto más refinado de reglas y estrategias para

guiar a los participantes en el desarrollo de comportamientos y competencias particulares y directamente transferibles al mundo real. Teniendo en cuenta los propósitos de cada uno de estos entornos interactivos, el desafío es lograr una unificación inteligente buscando combinar sus características a partir de los principios del aprendizaje constructivista en pos de contribuir a mejorar el proceso de enseñanza y aprendizaje de conceptos científicos o la comprensión de los fenómenos con un alto grado de abstracción. De esta manera, llamamos entorno interactivo educativo a aquellos entornos digitales que son desarrollados para ser utilizados en educación y que combinan las características de los mundos virtuales, los juegos y las simulaciones en pos de proporcionar un andamiaje para conceptualizar fenómenos dinámicos y permitir intercambios entre la cognición y los objetos simbólicos que representan el conocimiento científico. Entre las características que interesa conjugar en estos entornos interactivos educativos pueden mencionarse la optimización del sentido de lugar y presencia de los mundos virtuales, la participación y diversión por parte de los videojuegos, y el rigor y la transferencia de los aprendizajes para la resolución de nuevas situaciones de las simulaciones [Ald05].

Las tecnologías inmersivas proporcionan entornos alternativos de aprendizaje situado, favorecen las posibilidades de experimentar a través de la acción activa del usuario, de forma que éste pueda a su vez relacionarse con los elementos formativos e igualmente con el resto de elementos del entorno, volviéndose colaborativo y motivador [Daw14] [Man11]. Los ambientes inmersivos educativos abarcan diferentes medios (juegos, simulaciones, mundos virtuales, aprendizaje móvil, realidad aumentada, etc.). Pagano [Pag13] señala que el

aprendizaje inmersivo no es acerca de la tecnología, sino de los principios de diseño que permiten a los alumnos practicar en el contexto, aplicar sus conocimientos y mejorar sus habilidades y competencias.

Desde hace más de 10 años que se reconoce que los estudiantes están cambiando, son cada vez más pragmáticos, reclaman interacción y personalización, son muy visuales, a menudo son reacios a la lectura, quieren más material en menos tiempo, son muy computacionales. Es en este contexto que las simulaciones y los videojuegos, como entornos interactivos e inmersivos, cobran especial relevancia como herramientas educativas. Bouciguez et al. [Bou14] presenta un estudio de la población de estudiantes universitarios de ingeniería y sus preferencias en el uso de videojuegos, con el objetivo de estudiar la población destinataria de videojuego educativo para la enseñanza de ciencias.

Desarrollar entornos interactivos divertidos y atractivos pero que mantengan el potencial educativo para la formación en ciencias experimentales básicas plantea importantes desafíos. Bouciguez et al. [Bou13] describen algunos videojuegos educativos para enseñar Física y se caracterizan en función de la tecnología que involucra, los aportes y limitaciones al aprendizaje desde una perspectiva constructivista social. Esta caracterización puede brindar información útil sobre cómo utilizar un videojuego, qué requerimientos tecnológicos necesita y qué potencialidades puede aportar a una situación de enseñanza y aprendizaje.

Líneas de Investigación, Desarrollo e Innovación

- Aplicaciones de Realidad Aumentada en educación
- Aplicaciones de Realidad Aumentada en el ámbito militar

- Aplicaciones de Realidad Virtual con visualización de terrenos multiresolución
- Juegos educativos para la enseñanza de la física
- Interfaces avanzadas para discapacitados

Resultados y Objetivos

- Carrera de postgrado “Especialización en Computación Gráfica, Imágenes y Visión por Computadora”, Facultad de Informática de la UNLP. Dicha carrera el reconocimiento oficial provisorio por la CONEAU (Número de Carrera nueva N° 11.162/12).
- Dictado de un curso de doctorado conjunto entre docentes de la UNLP y de la UIB sobre la temática “Interfaces Avanzadas”
- Framework RAIOM en desarrollo, basado en realidad aumentada utilizando dispositivos móviles, visión por computador y sensores externos para el reconocimiento, detección, ubicación, identificación y suministro de información contextual
- Diseño e implementación de un algoritmo de visualización de terrenos 3D multiresolución utilizando la GPU [Bor15]
- Diseño e implementación de un sistema interactivo para discapacitados basado en la posición de las manos [Boj15]
- Se realizaron aplicaciones de prueba utilizando gafas de realidad aumentada “*video see-through*” Vuzix Wrap 920AR

Formación de recursos humanos

La formación de recursos humanos es prioritaria en esta línea, y por esto se implementó la carrera de postgrado mencionada en la sección anterior. En el marco de los proyectos conjuntos mencionados se ha podido contar con la visita de profesores de otras universidades nacionales y extranjeras (UNLP, UNS, UNICEN y UIB), con experiencia y formación en los temas propuestos, los cuales impartieron cursos de postgrado.

En la actualidad hay en curso diferentes tesis de postgrado en el marco de esta línea de investigación:

- María José Bouciguez “Ambientes virtuales altamente interactivos basados en videojuegos y simulaciones para la educación en ciencias” (tesis de doctorado en curso)

- Alejandro Mitaritonna. “Realidad Aumentada para la Identificación de Objetivos Militares” (tesis de doctorado en curso)

- Mario Vincenzi. “La Realidad Aumentada en la educación. Vigencia, proyecciones y límites”. (tesis de especialización en curso)

- Natalia Encina. “Evaluación de browsers de realidad aumentada para apoyar procesos de enseñanza - aprendizaje” (tesis de especialización en curso)

Referencias

[Ald05] Aldrich, C. (2005) *Learning by Doing: A Comprehensive Guide to Simulations, Computer Games, and Pedagogy in e-Learning and Other Educational Experiences*. John Wiley and Sons, ISBN N°:0-7879-7735-7

[Bal06] Balasubramanian, N. & Wilson, B. G. (2006). *Games and Simulations*, In C. Crawford et al. (Eds.), *ForeSITE* (<http://site.aace.org/pubs/foresite/>), Volume One, 2005. Proceedings of Society for Information Technology and Teacher

Education International Conference 2006. Chesapeake, VA: AACE.

[Boj15] Bojanovic, Filip; Manresa Yee, Cristina; Abásolo Guerrero, María José; Larrea, Martín Leonardo; Sanz, Cecilia Verónica (2015) *Get closer to activate it! An educational tool for people with multiple disabilities*. Proceedings of XXI Congreso Argentino de Ciencias de la Computación CACIC 2015. ISBN N°: 978-987-3724-37-4

[Bor15] Borrelli, Lucas; Abásolo Guerrero, María José (2015) *Visualización 3D de terrenos multiresolución basada en Shader Model 5* (2015) Proceedings of XXI Congreso Argentino de Ciencias de la Computación CACIC 2015. ISBN N°: 978-987-3724-37-4

[Bou13] María José Bouciguez, Graciela Santos, María José Abásolo (2013) *Potencialidad de los videojuegos en el aprendizaje de Física*. Actas de WEFA 2013 I Workshop de Enseñanza de Física en Argentina, Universidad Nacional del Centro de la Pcia. de Bs. As., ISBN 978-950-658-342-2.

[Bou14] María José Bouciguez, Graciela Santos. María José Abásolo *Towards the use of video games for learning: a survey about video games preferences of Engineering students*. Journal of Computer Science & Technology; vol. 14, no. 1, p. 25-31, ISSN 1666-6038, 2014 <http://journal.info.unlp.edu.ar/journal/journal38/papers/JCST-Apr14-4.pdf>

[Daw14] Dawley, L., & Dede. C. (2014) *Situated learning in virtual worlds and immersive simulations*. In J.M. Spector, M.D Merrill, J. Elen, & M.J. Bishop (Eds.), *The Handbook of Research for Educational Communications and Technology* (4th ed.). New York: Springer.

[Hew92] B. Hewett, C. Card, M. Gasen, Strong Perlman, and Verplank (1992) *Acm SIGCHI curricula for human-computer interaction*. <http://sigchi.org/cdg/index.html>

[ISO10] International Organization for Standardization (2010) *Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems*, 9241-210:2010

- [Mad11] Madden, Lester (2011) *Professional Augmented Reality Browsers for Smartphones: Programming for junaio, Layar and Wikitude*. Copyrighted Material. Wrox
- [Man11] Maniega, D., Yàñez, P. & Lara, P. (2011) *Lost In La Mancha: aprendizaje inmersivo online 3D*. Revista de comunicación y nuevas tecnologías. Revista ICONO 14. Año 9 Vol. 2. pp. 101-121. ISSN: 1697-8293. Madrid (España)
- [Man12a] Manresa-Yee, C.; Abásolo, M.J.; Mas Sansó, R.; Vénere, M.. (2011) *Realidad Virtual, Realidad Aumentada e Interfaces Basadas en Visión*. XV Escuela Internacional de Informática, XVII Congreso Argentino de Ciencia de la Computación CACIC 2011. Editorial EDULP, ISBN 978-950-34-0765-3
- [Man12b] C. Manresa-Yee, R. Mas, G. Moyà, M.J. Abásolo, J. Giacomantone (2012) *Interactive multi-sensory environment to control stereotypy behaviours*. Computer Science & Technology Series: XVII Argentina Congress of Computer Science (Selected Papers), pp.121-128, 978-950-34-0885-8
- [Man13] Cristina Manresa Yee, Joan Jordi Muntaner, and Diana Arellano (2013) *A motion-based interface to control environmental stimulation for children with severe to profound disabilities*. In CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13). ACM, New York, NY, USA, 7-12.
- [Man15] Cristina Manresa-Yee, Ann Morrison, and Joan Jordi Muntaner. 2015. First Insights with a Vibrotactile Interface for Children with Multiple Disabilities. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15). ACM, New York, NY, USA, 905-910.
- [Mit13] Mitaritonna, A.; Abásolo Guerrero, M. J. *Mejorando la conciencia situacional en operaciones militares utilizando la realidad aumentada* (2013) Proceedings of XVIII Congreso Argentino de Ciencias de la Computación. ISBN N°: 978-987-23963-1-2 pp. 356-365.
- [Mit15] Mitaritonna, A.; Abásolo Guerrero, M. J. (2015) *Improving Situational Awareness in Military Operations using Augmented Reality*. Proceedings of WSCG 2015. ISBN N°:978-80-86943-72-5, 2013.
- [Pag13] Pagano, K. O. (2013) *Immersive Learning: Designing for Authentic Practice*. Alexandria, VA: ASTD Press.
- [Squ08] Squire, K.D. (2008) *Game-based learning: An emerging paradigm for learning*. Performance Improvement Quarterly, 21 (2), 7-36. <http://www3.interscience.wiley.com/journal/120835177/issue>

Integridad de Datos. Evaluación basada en el método LSP

Estela Fritz¹; Germán Montejano^{1 2}; Pablo García¹; Silvia Bast¹;

¹Departamento de Matemática
Universidad Nacional de La Pampa
Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina
Tel.: +54-2954-245220 – Int. 7125
[fritzem, pablogarcia, silviabast]@exactas.unlpam.edu.ar

²Departamento de Informática
Universidad Nacional de San Luis
Ejército de los Andes 950 – (5700) San Luis – San Luis – Argentina
Tel.: +54-2652-424027 – Int. 251
gmonte@unsl.edu.ar – web: <http://www.unsl.edu.ar>

Resumen

En la actualidad, las Tecnologías de la Información y la Comunicación (TIC) tienen un rol fundamental en nuestra sociedad. El ámbito de la Educación no es la excepción en ello. Así, instancias y momentos del proceso educativo, han ido paulatinamente perdiendo su carácter presencial para dar paso a la virtualidad en la enseñanza. Este fenómeno abarca tanto las etapas iniciales de adquisición de conocimientos, como así también etapas posteriores como las producciones de los alumnos e incluso las instancias de evaluación y acreditación.

En este sentido, los Sistemas de Gestión de Aprendizaje (*LMS* por su nombre en inglés *Learning Management Systems*) han cobrado gran relevancia.

Considerando que un *LMS* con actividades en línea, supone múltiples puntos de acceso, acceso remoto e información compartida por numerosos usuarios, es de fundamental importancia garantizar la integridad de los datos.

Este y otros aspectos relevantes pueden ser analizados a través de la aplicación de diferentes métricas que tengan en cuenta

los estándares de ISO/IEC 25010 [7] e ISO/IEC 25012:2008[8], que definen un modelo general de calidad para datos almacenados de forma estructurada dentro de un sistema.

El presente trabajo sugiere algunos criterios en este sentido, proponiendo indicadores específicos para evaluar los LMS, con respecto a la integridad de los datos.

Palabras clave: Sistemas de Gestión de Aprendizaje, Integridad de los Datos, Seguridad de Sistemas de Información.

Contexto

Por Resolución N° 488/14 del Consejo Directivo de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa se acredita el Proyecto de Investigación: "Aspectos de Seguridad en Proyectos de Software", que se desarrolla en el ámbito de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa (UNLPam). El mismo es dirigido por el Doctor Germán Antonio Montejano y codirigido por el Magister

Pablo Marcelo García e incluye a la Licenciada Silvia Gabriela Bast y la Profesora Estela Marisa Fritz como investigadoras.

El presente trabajo se enmarca en el Proyecto de Investigación: "Aspectos de Seguridad en Proyectos de Software", que se desarrolla en el ámbito de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa y en el Proyecto de Investigación "Ingeniería de Software, Conceptos, Métodos y Herramientas en un Contexto de Ingeniería de Software en Evolución" de la Universidad Nacional de San Luis.

Introducción

La bibliografía que hace referencia al análisis de la calidad del software educativo, como por ejemplo [2] y [3], plantea este análisis tanto desde el punto de vista tecnológico como pedagógico.

En consecuencia, aspectos como corrección, fiabilidad, eficiencia, usabilidad, aparecen como los tópicos relevantes a medir. En este sentido existen trabajos que proponen métricas basadas fundamentalmente en encuestas al usuario y cuestionarios ([4]) cuyas respuestas se basan en categorías (bueno, muy bueno, aceptable, regular y otras) o tienen puntuaciones y así, realizan una evaluación técnica en el marco de los estándares de la Ingeniería del Software. También debe mencionarse que la utilización de los Sistemas de Gestión de Aprendizaje se ha diversificado. Ya no es sólo la gestión de recursos o un medio de comunicación entre el docente y los estudiantes (foros de discusión, cartelera de novedades, entre otros).

Es además un entorno donde el equipo de cátedra almacena información que es compartida por varios docentes, como por

ejemplo calificaciones parciales de los alumnos de un curso.

Asimismo hay tareas propuestas por el docente, con evaluación, cuya realización obligatoria es parte de la acreditación de los alumnos. Existen plazos para la presentación, establecidos de antemano. El sistema controla dichos plazos, no permitiendo la entrega fuera de término.

De lo expresado anteriormente cabe hacer una distinción:

El acceso a los materiales de un curso tales como apuntes y trabajos prácticos en formato digital, recursos multimediales y otros puede ser libre o restringido pudiendo estar éstos visibles o no visibles. Este aspecto está controlado por usuarios con roles específicos, como por ejemplo *profesor*. Existen otros roles como *estudiante* o *administrador*. Todos ellos poseen diferentes privilegios de acceso.

Así, de ese modo, un *profesor* restringe el acceso a un usuario *estudiante* a determinados materiales, como podría ser el enunciado de un examen, o la resolución del mismo en un tiempo previo a la instancia de evaluación. De la misma manera el *administrador* con sus privilegios puede realizar tareas de configuración del sistema y efectuar la gestión de los usuarios así como de los cursos y otros aspectos generales.

Para comprender la importancia de la integridad de los datos en las aplicaciones que administran, distribuyen y controlan actividades de enseñanza y aprendizaje, basta remitirse a algunas de las funciones de los Sistemas de Gestión de Aprendizaje, mencionadas en párrafos anteriores.

En el marco del proyecto de Investigación "Aspectos de seguridad en proyectos de software" se busca incorporar indicadores específicos de modo que la seguridad se convierta en un ítem de relevancia entre los criterios para la

evaluación de la calidad del software, principalmente en lo que se refiere a la integridad de los datos.

En [5] se expresa: “Desde la perspectiva de la seguridad de datos...la integridad de los datos es la garantía de que nadie pueda acceder a la información ni modificarla sin contar con la autorización necesaria.”

De este modo, cuando no se pueda garantizar la integridad de los datos, podrían ser modificados los permisos de usuario y los privilegios de acceso.

Esta situación permitiría a un *estudiante*, por ejemplo, asumir en el sistema el rol de *profesor*, y así, modificar sus calificaciones, o acceder a información que es no visible para sus permisos de acceso, entre otros.

En el mismo sentido un *profesor*, asumiendo el rol de *administrador* podría realizar cambios en la configuración del sistema que redunden en beneficio propio y eventualmente perjudiquen a otros usuarios. También, en los Sistemas de Gestión de Aprendizaje, los usuarios tienen diferentes privilegios y permisos de acceso a la información según su categoría (lectura, modificación), lo que incluye borrado, eliminación o destrucción de datos.

Este trabajo apunta a la propuesta de métricas e indicadores aplicados a la integridad de los datos en lo relacionado con la seguridad de la información.

Método LSP (*Logic Scoring of Preference*)

LSP (*Lógic Scoring of Preference*) es un método de evaluación multi-criterio que se basa en la definición de: 1) un Árbol de Criterios, 2) Criterios Elementales y 3) una Estructura de Agregación [6]. LSP es útil para analizar, comparar y seleccionar la mejor opción de un conjunto finito de alternativas (en el caso de este trabajo, la integridad de los datos en un LMS). LSP

consta de una serie de pasos que deben ejecutarse adecuadamente para llevar a cabo la evaluación .

La descripción del método en profundidad, según [12] excede los alcances del presente trabajo, pero a continuación, se describen brevemente los principales elementos:

1. Árbol de criterios

El Árbol de Criterios (AC) contiene las características deseables que los elementos a evaluar deben poseer. En el primer caso, se definen las características de alto nivel. A continuación, se descomponen en sub-características y así sucesivamente. El resultado de esta tarea es un árbol que contiene las principales características que los elementos, bajo evaluación, deben poseer.

2. Criterios elementales

En esta etapa se debe definir, por cada atributo (hoja) del AC, una función denominada Criterio Elemental (CE). Cada CE mapea el valor correspondiente al atributo en otro contenido en el intervalo [0,1] o [0,100]. Este valor se denomina Preferencia Elemental y representa el nivel de satisfacción correspondiente al atributo en cuestión, para el elemento evaluado.. Los CE se pueden clasificar como: *Absolutos* o *Relativos*

3. Estructura de agregación

Las Preferencias Elementales que resultan de aplicar los CE a los valores correspondientes a los atributos mensurables, deben agregarse a fin de obtener la Preferencia Global. Esta representa la satisfacción total del elemento evaluado, con respecto a todos los atributos definidos.

Para poder calcular la Preferencia Global, LSP cuenta con un conjunto de operadores lógicos (o funciones) propios del método. Todas las Preferencias se vuelven a agregar en el siguiente nivel de la Estructura de Agregación (EA).

4. Proceso de Evaluación

El proceso de evaluación se lleva a cabo asignando valores para cada atributo y cada elemento sometido a evaluación.

Líneas de Investigación, Desarrollo e Innovación

El grupo de trabajo investiga, básicamente sobre tres campos relacionados:

- Protección del anonimato de los votantes en sistemas de voto electrónico ([9]).
- Integridad de los datos de un sistema de e-Voting ([10]).
- Integridad de las bases de datos pertenecientes a un sistema de gestión de aprendizaje ([11]).

El presente trabajo corresponde a la tercera línea de investigación

Resultados y Objetivos

Una vez analizado el método LSP se definió el AC y la descomposición jerárquica hasta llegar a los CE para poder cuantificar el nivel de seguridad respecto a la integridad de los datos, aplicando para ello los operadores propios del método que permiten calcular la Preferencia global.

Algunos de los criterios propuestos para la construcción del AC son:

1. Nivel de seguridad en cuanto al acceso al sistema.
2. Restricciones en las operaciones de *upload* y *download* para las diversas categorías de usuarios.

Los criterios 1 y 2 pueden descomponerse en CE los cuales pueden mapear con valores en el intervalo [0,1] o [0,100] según la escala que se elija.

Algunos de los CE son:

- Número de usuarios que han mantenido derechos y privilegios de acceso históricos, en cada categoría.

- Registro de cuentas huérfanas o inactivas, cuantificado.
- Registro de accesos no autorizados a los datos, que hubiesen sido identificados.
- Cuantificación de accesos no autorizados en un período (en un mes/ en un año).

El cálculo de la Preferencia Global con los propios operadores, dará el marco para detectar fortalezas y debilidades del sistema en términos de la integridad de los datos.

A futuro, se prevé aplicar el AC elaborado, al LMS instalado en la Facultad de Ciencias Exactas y Naturales de la UNLPam.

Formación de Recursos Humanos

En el marco del presente proyecto se presentan los siguientes puntos relacionados con la formación de recursos humanos

- Pablo García realizó una estadía de un año en la Universidade Federal de Minas Gerais (UFMG), aprobando seminarios de posgrado y trabajando en el grupo “Criptografía Teórica y Aplicada”, dirigido por Jeroen van de Graaf, PhD.
- Silvia Bast está desarrollando su tesis para obtener el grado de “Especialista en Ingeniería de Software”. Su plan de trabajo fue aprobado y se planea su defensa para mayo de 2016. La tesis se titula: “Sistemas de E-Voting: Integridad de Datos” y está dirigida por el Dr. Germán Montejano (UNSL) y el Magister Pablo García (UNLPam).
- Pablo García defendió su tesis para obtener el grado de Magister en Ingeniería de Software de la Universidad Nacional de San Luis, bajo la dirección de Jeroen van de Graaf, PhD (UFMG) y Dr. Germán Montejano (UNSL). La tesis se tituló:

“Optimización de un Esquema Dining Cryptographers Asíncrono” y recibió la calificación sobresaliente.

- Silvia Bast y Pablo García completaron el cursado de la totalidad de los créditos exigidos en el Doctorado en Ingeniería Informática en la Facultad de Ciencias Físico Matemáticas y Naturales de la (UNSL).
- Estela Fritz está desarrollando su tesis para obtener el grado de “Especialista en Tecnologías Informáticas aplicadas en Educación”. Su plan de trabajo fue aprobado y se planea su defensa para octubre de 2016. La tesis se titula “Propuesta de clasificación de software libre utilizado en la enseñanza de la programación” y es dirigida por Mg. Alejandra Zangara (UNLP).

Referencias

- [1] **Uzal R., van de Graaf J., Montejano G., Riesco D., García P.**: “Inicio de la Línea de Investigación “Ingeniería de Software y Defensa Cibernética”. Memorias del XV Workshop de Investigadores en Ciencias de la Computación 2013 (WICC 2013).Ps.769 - 773. ISBN: 9789872817961. 2013.
- [2] **Gorga, G., Madoz, M., Pesado P.**, “*Hacia una propuesta de métrica para la evaluación de Software Educativo*”, CACIC 2000.Enlínea:<http://hdl.handle.net/10915/23514>
- [3] **Cataldi, Z.**, *Una Metodología para el Diseño, Desarrollo y Evaluación de Software Educativo*, Universidad Nacional de La Plata, Facultad de Informática, Tesis de Maestría, 2000.
- [4] **Squires,D., McDougall, A.**,*Cómo elegir y utilizar software educativo*, Ediciones Morata y Fundación Paideia, Primera Edición, 1997
- [5] **Gelbstein, E.** *La integridad de los datos: el aspecto más relegado de la seguridad de la información*, ISACA Journal, Volumen 6, 2011. En línea
- <http://www.isaca.org/Journal/Past-Issues/2011/Volume-6/Pages/Data-Integrity-Information-Securitys-Poor-Relation-spanish.aspx>.
- [6] J.J. Dujmovic. *A Method for Evaluation and Selection of Complex Hardware and Software Systems*. The 22nd Int'l Conference for the Resource Management and Performance Evaluation of Enterprise CS. CMG 96 Proceedings, 1:368–378, 1996.
- [7] **International Standard ISO/IEC 25010** Systems and software engineering Systems and software Quality Requirements and Evaluation (SQuaRE) System and software quality models. 2011
- [8] **International Standard. ISO/IEC 25012** Software Engineering – Software Quality Requirements and Evaluation (SQuaRE) - Data Quality Model. 2008
- [9] M. Claudia Albornoz, Enrique Miranda, Mario Berón, *Evaluación de Interfaces Gráficas de Usuario Usando LSP*. Universidad Nacional de San Luis, San Luis, Argentina
- [10] García P., Montejano G., Bast S., Fritz E.: “Seguridad Incondicional para el Anonimato en Sistemas de e-Voting”. XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015). 16 y 17 de abril de 2015. Facultad de Ciencias Exactas. Universidad Nacional de Salta. ISBN: 978-987-633-134-0. NACIONAL.
- [11]. Bast S., Montejano G., García P., Fritz E.: “Evaluación de la integridad de datos en Sistemas de e-Voting”. XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015). 16 y 17 de abril de 2015. Facultad de Ciencias Exactas. Universidad Nacional de Salta. ISBN:978-987-633-134-0. NACIONAL.
- [12] Fritz E., Montejano G., García P., Bast S.: “Integridad de Datos en Sistemas de Gestión de Aprendizaje”. XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015). 16 y 17 de abril de 2015. Facultad de Ciencias Exactas. Universidad Nacional de Salta. ISBN: 978-987-633-134-0. NACIONAL.

Desarrollo de Aplicaciones Móviles 3D Orientadas a Educación.

Pablo Thomas⁽¹⁾, Federico Cristina⁽¹⁾, Sebastián Dapoto⁽¹⁾, Patricia Pesado⁽¹⁾

⁽¹⁾Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – Universidad Nacional de La Plata

50 y 120 La Plata Buenos Aires

{pthomas, fcristina, sdapoto, ppesado}@lidi.info.unlp.edu.ar

Resumen

Se presenta una línea de investigación y desarrollo, que tiene por objeto estudiar temas relacionados con aspectos de Ingeniería de Software, orientados al desarrollo de aplicaciones móviles sobre diversas plataformas y entornos operativos. En particular, se pone el énfasis en el desarrollo de herramientas educativas para dispositivos móviles que mediante un entorno tridimensional favorezcan y simplifiquen el aprendizaje de los alumnos.

Palabras claves: Dispositivos Móviles - Aplicaciones 3D - Aplicaciones Multiplataforma - M-Learning

Contexto

Esta línea de Investigación forma parte del proyecto (2014-2018) “*Tecnologías para Sistemas de Software Distribuidos. Calidad en Sistemas y Procesos. Escenarios Educativos mediados por TICS*”, en particular del subproyecto *Ingeniería de Software en el Desarrollo de Aplicaciones para Dispositivos Móviles* del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por el Ministerio de Educación de la Nación.

Hay cooperación con Universidades de Argentina y se está trabajando con Universidades de Europa en proyectos financiados por el Ministerio de Ciencia y Tecnología de España y la AECID.

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática.

Por otra parte, se tiene financiamiento de Telefónica de Argentina en Becas de grado y posgrado.

Introducción

En la actualidad los dispositivos móviles forman parte de la vida cotidiana, tanto en el ámbito personal como en el profesional. Dichos dispositivos permiten a las personas interconectarse y acceder en todo momento y lugar a la información que necesitan.

Al mismo tiempo estos dispositivos son cada vez más sofisticados. La evolución tecnológica que han sufrido permite ejecutar aplicaciones cada vez más complejas y con exigentes requerimientos de hardware.

El ámbito educativo debe adaptarse a los cambios y nuevas formas de aprendizaje. Por ello debe considerarse la tecnología para incorporarla a los espacios de educación y usarla como facilitadora y agente de apoyo del aprendizaje.

M-learning (mobile learning) plantea métodos modernos de apoyo al proceso de aprendizaje mediante el uso de dispositivos móviles, tales como celulares, smartphones o tablets. El aprendizaje se transforma en una actividad personalizada, portátil, cooperativa e interactiva.

Una línea de la computación móvil de particular interés es su utilización con fines educativos. Los conceptos a incluir en los niveles iniciales de carreras informáticas y afines, brindan un escenario ideal para poder crear herramientas que soporten los pilares de m-learning.

Un caso de estudio en particular es la aplicación de escritorio R-Info, una herramienta de aprendizaje de los conceptos básicos para la construcción de algoritmos que se utiliza en los cursos de ingreso y asignaturas de primer año de la Facultad de Informática de la UNLP. Se trata de un robot virtual que ejecuta un conjunto de instrucciones sobre una ciudad. La utilización de R-Info sobre equipos móviles no era posible. Por este motivo surgió la necesidad de una nueva versión para este tipo de entorno.

Para desarrollar esta nueva herramienta se analizaron distintas alternativas de frameworks que posibiliten la generación de aplicaciones móviles multiplataforma. Una mejora adicional contemplada para esta nueva versión fue la de contar con un ambiente 3D de la ciudad y el robot, enriqueciendo la experiencia de aprendizaje y logrando un mayor interés por parte de los alumnos. Bajo estas consideraciones surgió el proyecto R-Info3D.

Otro caso de estudio fue el desarrollo de una aplicación móvil 3D que guía a los nuevos estudiantes dentro del edificio de la Facultad de Informática de la UNLP. El edificio provee la cartelería y las señalizaciones de forma adecuada, y dispone además de oficinas que orientan a los alumnos. No obstante, es importante disponer de una herramienta de software que sirva de apoyo para orientar a los estudiantes.

La herramienta móvil propuesta, denominada Recorrido Virtual 3D, consta de un modelo virtual tridimensional del edificio de la Facultad de Informática de la UNLP, en donde se presentan los puntos de interés más relevantes para los estudiantes ingresantes. De esta forma, dichos estudiantes pueden familiarizarse con la facultad y sus diferentes áreas aún antes de conocer físicamente el edificio.

Líneas de Investigación y Desarrollo

- Metodologías y Técnicas de la Ingeniería de Software y su aplicación en el

desarrollo de software para dispositivos móviles.

- Mobile Learning
- Aplicaciones Móviles Multiplataforma
- Aplicaciones Móviles 3D
- Frameworks para el desarrollo de Aplicaciones Móviles 3D
- Desarrollo de Ambientes Móviles para Iniciación en la Programación

Resultados esperados/obtenidos

Los resultados esperados/obtenidos se pueden resumir en:

- Avanzar en la capacitación continua de los miembros de la línea de investigación.
- Avanzar en el aprendizaje de frameworks que permiten desarrollar aplicaciones 3D multiplataforma, particularmente para dispositivos móviles.
- Avanzar en el desarrollo de aplicaciones educativas, teniendo como finalidad enriquecer las experiencias interactivas y motivar el aprendizaje mediante su uso.
- Se han evaluado diferentes herramientas para el desarrollo de aplicaciones móviles 3D. En particular se analizaron los frameworks Unity 3D y Unreal Engine, notando que difieren considerablemente en lo que respecta a lenguajes de programación, licencias de uso, soporte, comunidad, etc.
- Se ha desarrollado un prototipo móvil de una aplicación educativa en un ambiente tridimensional denominado R-Info3D, similar a la aplicación de escritorio R-Info, que puede ser utilizada desde distintas plataformas móviles. Dado que R-Info3D es un prototipo móvil, permite al alumno realizar pruebas de ejecución de algoritmos en cualquier momento y lugar. El entorno tridimensional de R-Info3D mejora la experiencia del alumno y permite una mejor visualización de las tareas que realiza el robot desde diferentes puntos de vista. Figura 1 y 2.

- Se pretende ampliar el prototipo móvil 3D desarrollado hasta alcanzar una funcionalidad similar a la de R-Info.



Figura 1. R-Info3D. Panel de edición. Vista en primera persona.

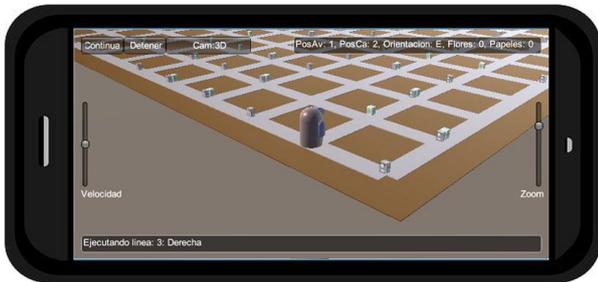


Figura 2. R-Info3D. Panel de ejecución. Vista en tercera persona.

- Se ha construido un modelo virtual 3D del edificio de la Facultad de Informática de la UNLP. Figura 3.
- Se ha desarrollado un prototipo móvil en un ambiente tridimensional denominado Recorrido Virtual 3D que sirve de guía para los estudiantes en sus primeras experiencias dentro de la facultad. Dicha herramienta les presenta a los estudiantes una serie de recorridos hacia los puntos de interés más relevantes dentro del edificio. Figura 4. Además, el prototipo permite recorrer el modelo virtual de forma libre, para poder conocer todas las áreas de la facultad, aún antes de visitar físicamente el edificio. Figura 5.

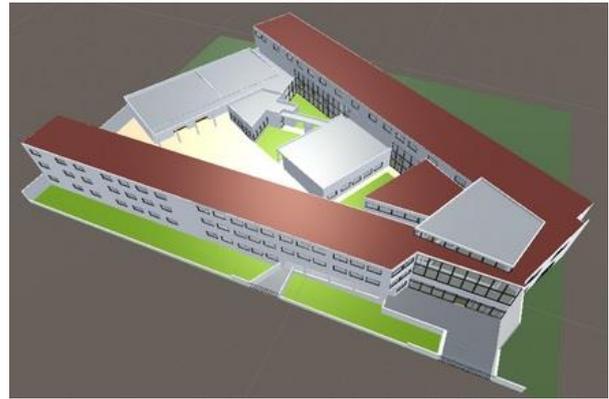


Figura 3. Modelo virtual 3D de la Facultad de Informática de la UNLP.



Figura 4. Recorrido Virtual 3D. Interface de la herramienta.



Figura 5. Recorrido Virtual 3D. Vista en el interior del edificio.

Formación de Recursos Humanos

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con el proyecto. Además participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

Referencias

1. Cristina, F.; Dapoto, S.; Thomas, P.; Pesado, P. "Prototipo móvil 3D para el aprendizaje de algoritmos básicos". XXI Congreso Argentino de Ciencias de la Computación CACIC 2015. Junín, Argentina. Octubre 2015. ISBN: 978-987-3724-37-4.
2. Cristina, F.; Dapoto, S.; Thomas, P.; Pesado, P. "A simplified multiplatform communication framework for mobile applications". IEEE International Conference on Computer Engineering & Systems (ICCES). December 2014. ISBN 978-1-4799-6593-9.
3. Material Didáctico del Curso de Ingreso 2014 - Facultad de Informática - UNLP. "Cap 2.2: Ambiente de programación del robot Rinfo"
<http://weblidi.info.unlp.edu.ar/catedras/ingreso/Material2014/EPA/GuiaEPA.pdf>.
4. De Giusti Armando, De Giusti Laura, Leibovich Fabiana, Sanchez Mariano, Rodriguez Eguren Sebastian. "Entorno interactivo multirrobot para el aprendizaje de conceptos de Concurrencia y Paralelismo". TE&ET 2014. Chilecito.
5. De Giusti Laura, Leibovich Fabiana, Sánchez Mariano, Rodriguez Eguren Sebastian, Chichizola Franco, Naiouf Marcelo, De Giusti Armando. "Herramienta interactiva para la enseñanza temprana de Concurrencia y Paralelismo: un caso de estudio". CACIC 2014. San Justo. ISBN: 978-987-3806-05-6.
6. De Giusti Laura, Leibovich Fabiana, Sánchez Mariano, Chichizola Franco, Naiouf Marcelo, De Giusti Armando. "Desafíos y herramientas para la enseñanza temprana de Concurrencia y Paralelismo". CACIC 2013. Mar del Plata. ISBN: 978-987-23963-1-2.
7. De Giusti Armando, Frati Emmanuel, Sanchez Mariano, De Giusti Laura. "LIDI Multi Robot Environment: Support software for concurrency learning in CS1". CTS 2012. ISBN: 978-1-4673-1380-3.
8. De Giusti Armando, Frati Emmanuel, Leibovich Fabiana, Sanchez Mariano, De Giusti Laura, Madoz María C. "LMRE: Un entorno multiprocesador para la enseñanza de conceptos de concurrencia en un curso CS1". CACIC 2011. La Plata. ISBN: 978-950-34-0756-1.
9. ICETI 2011. "Web based Interactive 3D Learning Objects for Learning Management Systems". Stefan Hesse, Stefan Gumhold. <https://www.inf.tu-dresden.de/content/institutes/smt/cg/publications/paper/ICETI2011.pdf>.
10. Kantel E., Tovar G., Serrano A. "Diseño de un Entorno Colaborativo Móvil para Apoyo al Aprendizaje a través de Dispositivos Móviles de Tercera Generación." IEEE-RITA 5, no. 4 (2010): 146-151.
11. 2009 LA Web Congress. "Interacting with 3D Learning Objects". Rosa Paredes, J. Alfredo Sánchez, Liliana Rojas, Daniel Strazzulla, Ronel Martínez-Teutle. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5341602>.
12. Cruz e Costa J., Ojala T., Korhonen J. "Mobile lecture interaction: Making Technology and learning click". IADIS International Conference Mobile Learning 2008.
13. Unity 3D Homepage: <https://unity3d.com/>.
14. Unreal Engine Homepage: <https://www.unrealengine.com/>.

Aplicando nuevos aspectos en la Programación de Computadoras

Lidia Lopez, Silvia Amaro

{lidia.lopez, silvia.amaro,}@fi.uncoma.edu.ar

Ana Alonso, Ingrid Godoy, Marcela Leiva, Juan Cruz Piñero

{ana.alonso, ingrid.godoy, marcela.leiva, jpinero}@fi.uncoma.edu.ar

Departamento de Programación - Facultad de Informática

Universidad del Comahue

Buenos Aires 1400, Q8300IBX Neuquén, Argentina

Resumen

En la actualidad, programar en computación no sólo se refiere a las computadoras, sino que ha sumado otros dispositivos y otras arquitecturas. Enseñar a programar a partir de pequeños proyectos que utilizan nuevas tecnologías ofrece otra perspectiva para la motivación. El trabajo colaborativo ha cobrado importancia, de manera que los programadores trabajan en espacios virtuales construyendo partes de aplicaciones que luego resultan en implementaciones globales.

Desde nuestro proyecto seguimos enfocados en generar propuestas diferentes, que resulten atractivas, de manera de aportar hábitos en el aprendizaje de la programación y el desarrollo de aplicaciones, aumentando el grado de interacción de los estudiantes entre sí, la motivación, las habilidades de gestión de la información y la construcción del conocimiento.

En los dos primeros años hemos trabajado en adoptar diferentes enfoques que prioricen las prácticas de aprendizaje colaborativo, y aprendizaje basado en proyectos competitivo. En la búsqueda de estrategias para la enseñanza de programación surgen los torneos utilizando herramientas tecnológicas para su evaluación y rankeado. Durante este último año hemos incorporado en la enseñanza de la programación, la construcción de pequeños proyectos de programación para móviles y para ambientes concurrentes.

Palabras clave: programación de computadoras, evolución de la programación, enseñanza de la programación, programación móvil, programación concurrente, programación competitiva .

Contexto

Esta investigación se enmarca dentro del proyecto "Aspectos Avanzados en Programación de Computadoras" aprobado y financiado por la Universidad Nacional del Comahue desde el año 2013, tiene como objetivo general "Establecer modelos, procesos y técnicas de programación de computadoras cuyo alcance abarca desde la enseñanza de la programación a partir de las TICs, hasta aplicaciones específicas que complementen la enseñanza y la construcción de software.

Asimismo esta vinculado con los proyectos de extensión *Entrenamiento en la programación de la computadora a partir de una aplicación para competencias* [12] y *Olimpiadas de Programación para escuela media* [14], ejecutados durante los años 2014 y 2015 respectivamente y *La Competencia como eje motivador para acercarnos a la Programación* que esta en ejecución el corriente año.

Introducción

Hemos trabajado en el estudio y análisis de estrategias con el objeto de establecer el enfoque metodológico para la resolución de problemas, la construcción algorítmica y la programación de alto nivel, buscando mejorar la motivación de los estudiantes [1, 2, 5].

Enseñar programación significa enseñar la tecnología y la ciencia. [15]. La tecnología consiste de herramientas, técnicas y estándares que permiten hacer programación. La ciencia consiste de una teoría amplia y profunda que permite entender la programación. Entonces conocer las herramientas y los conceptos prepara al estudiante para el presente y para los desafíos futuros. que le presentará la evolución tecnológica.

Las Tecnologías de la Información y de las Comunicaciones (TIC) forman parte activa de nuestra sociedad. Las organizaciones confían partes fundamentales de sus mecanismos de producción y gestión a la informática, y a ésta se le exigen altos niveles de confianza y robustez

Al mismo tiempo el panorama es dinámico, hay que estar preparados para incorporar las innovaciones necesarias para prestar los servicios que se requieren en la llamada Sociedad del Conocimiento, sin perder los niveles de confianza alcanzados. Los profesionales deben ser capaces de gestionar y adaptar los sistemas, tomar decisiones influyentes y de trabajar en equipo para el logro de los objetivos comunes. En la actualidad los desafíos tecnológicos son muchos y se producen avances continuamente: la "Internet de las cosas" [20], el desarrollo de "Aplicaciones Reactivas" [21], entre otros.

El estudio de ciertos tópicos tales como refactorización de código [8, 10], programación multiparadigma [17], metaprogramación [9], programación extrema, lenguajes específicos de dominio [7], implica que se conozcan los conceptos básicos pero además

que se diseñen prácticas y mecanismos para ayudar al estudiante a comprenderlos motivándolos con desarrollos que les resulten atractivos. En este escenario la programación web y el desarrollo de aplicaciones móviles logra gran interés por parte del alumnado

Muchas veces los estudiantes no se sienten preparados para programar aunque hayan aprobado las asignaturas respectivas [6], ya que además de desarrollarse en esta habilidad, se pretende que desarrolle otras: que el alumno sea capaz de seguir capacitándose en herramientas, técnicas, paradigmas y lenguajes de programación a medida que lo necesita, hacer uso de las TICs en el tiempo que prefiera para capacitarse, poder emplear la tecnología para comunicarse con otros y aprovechar los aportes en las comunidades relacionadas con la programación y desarrollo de aplicaciones e incluso hacer sus aportes allí [1, 2, 3, 4].

Por otro lado la propuesta de trabajar en base a competencias de programación intenta lograr por parte de los estudiantes un trabajo colaborativo dentro de un ambiente competitivo. Las competencias requieren que los estudiantes se organicen y asignen roles dentro del equipo. De esta manera ellos pueden aprender de sus puntos de vista, dar y recibir ayuda de sus compañeros de clase y ayudarse mutuamente para investigar de manera más profunda acerca de lo que están aprendiendo.

Líneas de Investigación, Desarrollo e Innovación

Las líneas de investigación que se identifican dentro del proyecto no son distantes sino que están conectadas transversalmente.

Por un lado la enseñanza de la Programación que se orienta a la incorporación de TICs para crear espacios

virtuales de comunicación interactiva en la que los estudiantes puedan intervenir. Junto con esto la incorporación del uso de diversas aplicaciones Open Source para la enseñanza inicial de los conceptos de programación, y aún para la enseñanza de conceptos más avanzados, que será además un factor de motivación y facilitará la comprensión de los conceptos.

Por otro lado el diseño y desarrollo de aplicaciones específicas que asistan en la enseñanza de la programación. Para ello se considerarán enfoques avanzados de programación, según las siguientes líneas de investigación:

- Programación Multiparadigma:
¿Cómo enseñar programación sin estar atado por las limitaciones de las herramientas y lenguajes existentes?. La respuesta inicial a esta pregunta se encuentra en la propuesta de Peter Van Roy [15, 16] respecto a considerar un enfoque basado en conceptos que logra la independencia deseada y la robustez tanto en la teoría como en la práctica. Los conceptos y construcciones de dos o más paradigmas se integran dando origen a un ambiente de desarrollo multiparadigma, en el que será posible lograr mejores soluciones a los problemas, desarrollando aplicaciones aplicando el paradigma más adecuado a cada subdominio [17]
- Programación Extrema y Refactorización:
Hemos creado, en conjunto con la Subsecretaría de TI de la Universidad, el grupo de desarrollo Euclides euclides.uncoma.edu.ar buscando agrupar y trabajar colaborativamente con estudiantes de asignaturas avanzadas en la construcción de programas a partir de metodologías ágiles aplicando programación extrema. Los desarrollos se basan en testing y refactorización continua. Se trata de la evolución del código que está en funcionamiento de manera eficiente [8, 10]

- Programación Competitiva:
Se está trabajando sobre una herramienta para competencias de programación desarrollada en la Facultad de Informática de la Universidad del Comahue: HORNERO [19]. A partir de un estudio comparativo de las herramientas existentes para la programación competitiva y tomando como punto de partida la competencia de la ACM-ICPC, se pretende generar una nueva versión de HORNERO, de manera que pueda ser utilizada como *objeto de aprendizaje* [18, 23] por los distintos actores involucrados en la enseñanza y aprendizaje de la programación.

Resultados y Objetivos

Nuestro proyecto apunta al estudio de técnicas de enseñanza más actuales, que incorporen el uso de las TIC y otras herramientas digitales.

A partir de la actualización de los contenidos de los planes de estudio de las carreras de informática, hemos ido implementando nuevos espacios teórico-prácticos para desarrollar la construcción de programas en entornos concurrentes y para tecnología móvil.

Por otro lado, hemos adoptado diferentes enfoques priorizando prácticas de aprendizaje colaborativo, basado en proyectos y aprendizaje competitivo, enfocados en la motivación a partir de aspectos lúdicos, definiendo claramente los criterios de evaluación.

Las experiencias realizadas han permitido gestionar el desarrollo y ejecución de ejercicios de programación dentro de dos contextos: el curricular y otro de aprendizaje no formal (precalentamientos, competencias de programación).

El desarrollo de torneos de

programación provee una forma de fomentar la creatividad, el trabajo en equipo y la innovación en la construcción de nuevos programas de software permitiendo al mismo tiempo a los estudiantes probar su habilidad para trabajar bajo presión. Esta divulgación lleva a que el destinatario asocie el conocimiento con aplicaciones que él mismo puede construir y modificar, y favorece el pensamiento reflexivo y crítico.

El objetivo que perseguimos incluye la búsqueda de estrategias para mejorar la enseñanza y aprendizaje de programación. Esto requiere el establecimiento de modelos, procesos, técnicas y herramientas orientadas a complementar la enseñanza. En este contexto pensar a la programación con una visión multiparadigma ó poliparadigma contribuye a lograr un aprendizaje de programación que sea independiente de un lenguaje en particular y prepara al estudiante para adaptarse a los distintos e incluso nuevos lenguajes de programación.

La propuesta no busca sustituir los sistemas de aprendizaje tradicionales, sino complementarlos, reforzando el aprendizaje de los estudiantes, debido a que relaciona el conocimiento con aplicaciones de la realidad, y contribuye a que los jóvenes se interesen en abordar áreas científicas para un futuro profesional, que le presentará grandes desafíos.

Formación de Recursos Humanos

El equipo de investigación está conformado por 2 Profesoras, 3 Asistentes de Docencia y 1 Ayudante de Primera. Inicialmente comenzamos con 2 estudiantes. Uno de ellos ha terminado su tesis de grado en el marco del proyecto de investigación [17].

Relacionados al proyecto se cuenta con 2 tesis de posgrado y 1 tesina de grado en curso. La tesina de grado corresponde a una beca TICs de la convocatoria 2012 desde la Agencia Nacional de Promoción Científica y Tecnológica, a través del FONSOFT, en el

marco del Programa para promover la innovación productiva a través del fortalecimiento y consolidación de Capital Humano aplicado a la industria de las Tecnologías de la Información y las Comunicaciones

Referencias

- [1] *Tutorías virtuales: acortando distancias a través de la comunicación electrónica*. J.Fernández Pinto. Quaderns Digitals 29.
- [2] *Educación a distancia, educación presencial y usos de la tecnología: una tríada para el progreso educativo*. Albert Sangra Morer. Edutec 15. Mayo 2002.
- [3] *Moodle*. <http://moodle.org.ar>
- [4] *PEDCO (Plataforma de Educación a Distancia Universidad Nacional del Comahue)*. C. Fracchia, A. Alonso de Armiño. Workshop de Tecnología Informática aplicada en Educación (WTIE). Congreso Argentino de Ciencias de la Computación. Universidad Nacional de La Matanza, San Justo, Buenos Aires, Octubre 2004.
- [5] *Moodle en la enseñanza universitaria: uso novedoso de la actividad libro*, Fracchia, C., Alonso de Armiño Ana C. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología TEYET 10. 2013.
- [6] *Reflections on the Teaching of Programming: Methods and Implementations*, Bennedsen, Jens. , 2008, Springer.
- [7] *DSLs In Action*, Ghost, Debasish. 2011, Manning Publications Co. ISBN: 978-1-935182-45-0
- [8] *Refactoring to Patterns*, Kerievsky, Joshua. 2011, Addison-Wesley. ISBN: 0-321-21335-1
- [9] *Metaprogramming Ruby*, Perrotta, Paolo. 2010, Jill Steimberg. ISBN: 978-1-934356-47-0
- [10] *Proceedings of the Fourth Workshop on Refactoring tools, WRT 2011*, Dig,

Danny and Batory, Don. 2011. ISBN: 978-1-4503-0579-2.

[11] *Aprenda a Pensar como un programador*, Downey Allan, Elkner Jeffrey, Meyers Chris. 2002 . ISBN: 0-9716775-0-6

[12] *Entrenamiento en la programación de la computadora a partir de una aplicación para competencias*, Proyecto de Extensión Universidad Nacional del Comahue, 2014, Directora: Ingrid Godoy.

[13] *Realización de Torneos de programación como estrategia para la enseñanza y el aprendizaje de programación*, Fracchia, Claudia Carina; Kogan, Pablo; Alonso de Armiño, Ana C.; Godoy, Ingrid; López, Lidia Marina. XX Congreso Argentino de Ciencias de la Computación. Buenos Aires, 2014. Pág 42-51. ISBN 978-987-3806-05-6.

[14] *Olimpiadas de Programación para Escuela Media*, Proyecto de Extensión Universidad Nacional del Comahue, 2015, Directora: Silvia Amaro.

[15] *Concepts, Techniques and Models of Computer Programming*, Peter Van Roy and Serif Haridi, The MIT Press, 2004. ISBN 13:978-0262220699

[16] *The CTM Approach for Teaching and Learning Programming*, Peter Van Roy, Horizons in Computer Science Research. Vol 2. Cap 5. Thomas S. Clary Ed. 2011. ISBN 978-1-61761-439-2

[17] *Análisis Comparativo de Ambientes de*

Programación Multiparadigma, Yanina Paola Perez. Tesina de Licenciatura en Ciencias de la Computación. Universidad Nacional del Comahue. 2015.

[18] *Learning Objects: A Practical Definition*, Rory McGreal, International Journal of Instructional Technology & Distance Learning., vol 1, nro 9. Sept 2004. ISSN 1550-6908

[19] *Hornero: Aplicación para Gestión de Torneos de Programación, Multi-Paradigma, Multi-Plataforma, Multi-Lenguaje*, Carina Frachia, Silvia Amaro, Pablo Kogan, III Jornadas de Cloud Computing and Big Data JCC&BD 2015. Universidad Nacional de La Plata. Julio 2015.

[20] *Building the Web of Things*. Dominique D. Guinard and Vlad M. Trifa. Manning Publications 2015. ISBN 9781617292682.

[21] *Reactive Web Applications, with Scala, Play, Akka, and Reactive Streams*. Manuel Bernhardt. Manning Publications 2016. ISBN 9781633430099

[22] *Guide to Teaching Computer Science: An Activity Based Approach*, Orit Hazzan, Tami Lapidat, Noa Ragonis. 2011, Springer. ISBN: 978-0-857294-42-5

[23] *Connectiong learning Objects to instructional design theory: A definition, a metaphor, and a taxonomy*. David Wiley. Digital Learning Environments Research Group.

TIC Sostenibles para la educación y concienciación

Javier Díaz¹, Viviana M. Ambrosi^{1,2}, Nestor Castro¹, Damián Candia¹, Edgar F. Vega¹, Anahí S. Rodríguez¹

1- LINTI – Facultad de Informática – UNLP

2- Comisión de Investigaciones Científicas de la Provincia de Buenos Aires - CIC

javier.diaz@info.unlp.edu.ar, vambrosi@info.unlp.edu.ar, ncastro@isis.unlp.edu.ar, dcandia@linti.unlp.edu.ar, evega@linti.unlp.edu.ar, arodriguez@linti.unlp.edu.ar.

Resumen

El vertiginoso avance de las TICs a nivel mundial ha generado un incremento en los niveles de producción y descarte tanto de equipos como componentes. Esto ha provocado un incremento exponencial del equipamiento informático en desuso u obsoleto que se desecha a nivel mundial, cuyo destino final en muchos casos son depósitos y en otros basurales a cielo abierto.

Pero el problema no es sólo una cuestión de espacios. Existe un problema aún mayor, de carácter educativo-cultural, debido al desconocimiento de la población sobre los graves trastornos a la salud y perjuicios al medioambiente generados al ser desechados o tratados incorrectamente.

Es nuestra obligación como docentes/investigadores de una Universidad pública contribuir a una educación de calidad dando soluciones concretas, acorde con los requerimientos individuales-sociales, útil para la inserción social, y extensible hacia los sectores desfavorecidos de la sociedad.

Pero sumando la participación de alumnos y docentes.

Para ello desde el año 2009 la Facultad de Informática de la Universidad Nacional de La Plata y el LINTI llevan adelante el Proyecto E-Basura, brindando una solución interdisciplinaria a la problemática de los Residuos de Aparatos Eléctricos y Electrónicos (RAEE). Logrando una integración I+D+I.

Palabras clave: RAEE, Basura Electrónica, E-Basura, Tecnologías Verdes, Desarrollo Sostenible, Green IT

Contexto

El programa E-Basura [1] perteneciente al LINTI (Laboratorio en Investigación en Nuevas Tecnologías Informáticas) [2] de la Universidad Nacional de la Plata es un proyecto de extensión universitaria.

El mismo interactúa con otros Proyectos de Extensión de la Facultad de Informática [3] y se vincula con otras unidades académicas para ampliar y

potenciar su alcance, en formación, concienciación, marco legal, cuidado ambiental y reaprovechamiento en la cadena de valor de los RAEE.

Durante el año 2015 se trabajó en conjunto con las unidades académicas de Ciencias Jurídicas y Sociales, Periodismo y Comunicación Social, y Humanidades y Ciencias de la Educación.

Introducción

La contaminación que provoca la basura electrónica y su toxicidad para la salud se ha tornado un problema para todos los sectores de la comunidad y una gran preocupación para organismos nacionales e internacionales tales como Gobiernos [4], ONGs y Universidades, quienes han comenzado a establecer un compromiso serio para intervenir sobre esta cuestión.

En América-Latina, Argentina es considerada uno de los países líderes en cuanto a su evolución en el sector de las TICs según lo que se observa en [5]. En cuanto a los niveles de descarte, la Universidad de las Naciones Unidas y el Instituto de Estudio Avanzado de la Sostenibilidad IAS, se estiman que en 2014 se produjeron alrededor de 300.000 Toneladas de residuos electrónicos, lo cual equivale a 7 Kg. por habitante por año. [6]

Ante esta problemática el objetivo de nuestro programa fue transformar un problema en beneficios y oportunidades teniendo en cuenta que:

- Muchos equipos informáticos reacondicionados en forma adecuada pueden ser donados. Reduciendo por un lado los niveles de basura electrónica, y por otro la brecha social y digital.
- La extensión del ciclo de vida de los equipos reduce el consumo de

materiales y energía generados durante la fabricación de una nueva computadora. No siempre se requiere contar con equipamiento de última generación para acceder a la tecnología y a la alfabetización digital.

- Aquello que no se reusa/recicla termina siendo un residuo peligroso con destino final incierto que compromete la salud y el ambiente.
- Una de las principales prioridades debe ser mejorar la “calidad” y “equidad” de la educación en todos los niveles
- La educación es un instrumento para la formación integral de los jóvenes.
- La capacitación en oficios de jóvenes y adultos mayores incrementa la autoestima y mejora la inserción laboral y la tecnología puede utilizarse para ello.
- Los alumnos deben ser preparados de acuerdo a los avances de la tecnología pero sin olvidar el medio ambiente.

De esta forma el programa E-Basura [1] implementó una plataforma que atiende la problemática de los RAEE (Residuos de Aparatos Eléctricos y Electrónicos) desde 2009 en Argentina.

Acciones que realiza el programa

Educación y Concientización

El primer paso para poder llevar a la práctica acciones que aporten al bienestar ecológico, se relaciona con el conocimiento pleno de los riesgos que generan los basureros electrónicos. Para ello, es necesario dialogar sobre el medio ambiente y la responsabilidad social que a cada sujeto le corresponde individualmente.

Desde E-Basura [1] se propician encuentros con instituciones y alumnos de

todos los niveles, siendo los niños de los colegios primarios de la zona quienes se constituyen como los primeros interlocutores, y por su intermedio se llega a familiares, amigos, conocidos y al resto de la comunidad.

Prolongar la vida útil del equipamiento informático

Extender la “vida útil” de los equipos electrónicos, es una de las premisas del proyecto. Esto significa que si un equipamiento es desechado por obsoleto, los alumnos y profesionales que trabajan en el mismo los recuperan para su uso, extendiendo su vida útil sin convertirse en basura tecnológica, disminuyendo los niveles de la misma y cuidando los recursos naturales.

Reutilizar

La reutilización y recuperación de partes de una computadora es de vital importancia para el proyecto. Los discos son recursos escasos y en su destrucción física no se contempla el principio de “reutilización”. Para asegurar la confidencialidad se utilizan herramientas de software para el borrado seguro de la información contenida en los discos.

Donar

Todo el material recuperado es donado a comedores populares, escuelas de nivel inicial, media, técnica, rural, especial, bibliotecas, ONG, OSC, organizaciones de la tercera edad, hospitales y dispensarios públicos, entre otras organizaciones. El objetivo es contribuir a la alfabetización digital y a la reducción de la brecha digital.

Software Libre

El proyecto promueve el uso de Software Libre. Los equipos son instalados con Lihuen-GNU Linux [7] una distribución propia de la Facultad de Informática [3] que incluye además aplicaciones educativas.

Disposición final segura

Todo aquel material que no puede ser reutilizado es retirado por empresas que cuentan con las respectivas certificaciones ambientales (Provinciales y Nacionales).

Escuela de Oficios y Pasantías

Con el objetivo de contribuir a la equidad social se dictan cursos de oficios en reparación de computadoras de forma gratuita a diversos grupos, posibilitándoles una salida laboral e inserción social.

Formación de alumnos universitarios

Un punto destacado de la iniciativa, que cierra el ciclo virtuoso, es la formación de los alumnos universitarios en el cuidado ambiental, la ayuda y el respeto por el prójimo.

Un problema transformado en inclusión digital y social

Transformar un problema en oportunidades es todo un desafío, y que además logre convertirse en inclusión digital y social para los sectores más desfavorecidos de la sociedad con el aporte de estudiantes universitarios, es una puesta en valor a la iniciativa.

Líneas de Investigación, Desarrollo e Innovación

Desde el año 2007 el LINTI comenzó a investigar sobre el uso de las TIC en relación a los temas de salud, inclusión y educación desde su proyecto “Redes, Seguridad y Desarrollo de Aplicaciones para e-Educación, e-Salud, e-Gobierno y E-Inclusión. A partir del 2008 comenzó a investigar sobre la problemática de los RAEE.

Como consecuencia de dichas investigaciones surge en el año 2009 el proyecto E-Basura, el cual tiene como objetivo extender la vida útil de los equipos informáticos y reducir la brecha digital que existe en la sociedad mediante

la donación del equipamiento informático recuperado y la formación.

Dentro del proyecto trabajan en conjunto tanto alumnos como docentes de la Facultad de Informática [3] de la UNLP [8], esto permite formar a los alumnos tanto en investigación como en extensión y su vinculación con la docencia.

Como resultado de las investigaciones se determinó que los residuos electrónicos son el eslabón final en el segmento de las TICs, y que se debía tener una visión global de la tecnologías verdes. De allí surge la necesidad de incorporar una materia específica "Green IT" para ser dictada en el quinto año de las carreras de las Licenciaturas en Sistemas y en Informática de la Facultad de Informática. Materia de carácter innovador dentro de los planes de estudio de las Universidades de Latinoamérica [9].

Sumado a esto, un convenio firmado entre la UIT [10] y las Universidades Argentinas permitió enriquecer la labor y generar contribuciones para las buenas prácticas en el uso de las TICs y el cuidado ambiental, y su aplicación en las actividades de I+D+E.

Resultados y Objetivos

El proyecto se refuerza el compromiso de la Universidad Pública con la Sociedad, persiguiendo los siguientes objetivos:

- Generar conciencia y educar en el ámbito de la UNLP y en el resto del país acerca del riesgo que conlleva la falta de tratamiento de los RAEE.
- Propiciar la reutilización tecnológica prolongando el tiempo de vida del equipamiento informático a través de su restauración y puesta a punto.
- Posibilitar el acceso a la tecnología, a través de la donación de equipos informáticos a los sectores más desfavorecidos de la sociedad

(instituciones, organizaciones sociales, comedores populares, escuelas, bibliotecas y demás entidades de bien público y sin fines de lucro) para reducir la brecha digital y contribuir a la equidad social.

- Contribuir a la disposición final segura, evitando el estadio final en quemas y basurales.
- Generar acciones activas y alianzas con empresas que permitan minimizar el impacto ecológico de los RAEE.
- Fomentar la Responsabilidad Social Corporativa buscando alianzas que permitan potenciar la iniciativa en post del beneficio social desde la Empresa, el Estado y la Universidad.

Resultados del 2009 a la fecha:

- 62.511 alumnos/personas beneficiadas con posibilidades de alfabetización digital luego de la donación de equipamiento informático
- Participación en eventos y jornadas con el objetivo del acercamiento a la sociedad. El proyecto ha participado en 178 eventos de diverso tipo
- Ha realizado más de 185 notas en medios periodísticos.
- Se capacitaron a 561 alumnos en el oficio de reparación y armado de computadoras, favoreciendo su inserción laboral.

Ha recibido varias distinciones:

- 1er Premio "Inclusión Digital 2009" de Programa Nacional para la Sociedad de la Información (PSI) [11] dependiente de la Secretaría de Comunicaciones de Nación, en la categoría "C" para iniciativas no gubernamentales y/o de extensión universitaria que realicen actividades vinculadas a la reducción de la Brecha Digital.
- Declarado de Interés Provincial por el Gobierno de la Provincia de Buenos Aires (2012)

- 1er Proyecto ganador del Concurso del Banco Mundial. Seleccionado entre 30 países. Proyecto codirigido junto a jóvenes empleados del Banco Mundial para la capacitación técnica de jóvenes con el fin de mejorar empleabilidad e inclusión social. (2012)
- Firma de convenio de colaboración institucional con el gobierno de la provincia de Buenos Aires y cesión del predio (2015).
- Subsidiado en las Convocatorias a Proyectos de Extensión Universitaria desde el año 2010 a la fecha.
- Incluido en el Plan Estratégico de la UNLP (2010-2018).
- Seleccionado por el Proyecto Econormas Mercosur para la Promoción de la Producción y Consumo Sostenible y las Buenas Prácticas Ambientales. (2014)
- Marca Registrada (2013)
- Mención de Econormas Mercosur, por haber participado en la implementación de buenas prácticas de gestión ambiental y producción más limpia desarrollada con las PyMES del sector RAEE en Argentina (2014).
- Declarado “Programa E-Basura e Informática Verde” a partir del año 2014 en la Facultad de Informática.

Formación de Recursos Humanos

La problemática del cuidado del medio ambiente ha cobrado gran importancia en los últimos años a nivel mundial. Donde la solución parte de la educación en todos los niveles.

Por ello desde el año 2009 se trabaja sobre temáticas sociales y ambientales, junto a alumnos de distintas carreras universitarias de la UNLP. Este tipo de actividades habilita nuevas competencias que son muy valoradas como parte de su formación profesional.

Complementa la formación con la educación, sensibilización ambiental, y la Escuela de Oficio en el armado y reparación de PC para dar una salida laboral, fomentando la reutilización, la protección ambiental y la ayuda social.

Por otro lado, la relación entre I+D+E ha permitido integrar la temática y la formación de recursos humanos en la parte de investigación, en la docencia incorporando conceptos de Green IT a la currícula de las Licenciaturas en Informática y Sistemas. Ha favorecido la interacción entre alumnos, docentes e investigadores formados, y la realización de tesis y trabajos de grado para diferentes ramas de las ciencias.

Referencias

- 1- <http://e-basura.linti.unlp.edu.ar>
- 2- <http://www.linti.unlp.edu.ar>
- 3- <http://www.info.unlp.edu.ar>
- 4- <http://www.basel.int>
- 5 - Banco de Desarrollo de América Latina, “Hacia la transformación digital de América Latina: las infraestructuras y los servicios TIC en la región”
http://publicaciones.caf.com/media/39809/informe_tecnologiacaf.pdf
- 6- UNU-IAS, “eWaste en América Latina. Análisis estadístico y recomendaciones de política pública, noviembre 2015”

www.gsma.com/latinamerica/wp-content/uploads/2015/11/gsma-unu-ewaste2015-spa.pdf

7-<http://lihuen.linti.unlp.edu.ar/index.php>

8- <http://www.unlp.edu.ar>

9-Díaz, Javier F., Ambrosi, Viviana, Castro, Néstor, Banchoff Tzancoff, Claudia, M. Raimundo, Marcelo, “Porque incluir Green IT en la currícula de Informática”, TE & ET 2014.

10- www.itu.int/es/Pages/default.aspx

11- Programa Nacional para la Sociedad de la Información (PSI),
<http://www.psi.gob.ar/>.
http://www.psi.gob.ar/civitas/modulos/noticias/adjuntos/adjunto_13253_2.pdf

PROGRAMAR en la Escuela

Claudia Queiruga, Claudia Banchoff, Sofía Martín, Vanesa Aybar Rosales,
Fernando López

LINTI. Facultad de Informática. Universidad Nacional de La Plata
50 y 120. La Plata
{claudiaq, cbanchoff, [vaybar](mailto:vaybar@info.unlp.edu.ar)}@info.unlp.edu.ar, {smartin,
flopez}@linti.unlp.edu.ar

Resumen

Enseñar a programar en la escuela es hoy una preocupación global. Numerosas actividades e iniciativas dan cuenta de ello.

Hace unos años, se hablaba de brecha digital refiriéndose a aquellos que tenían acceso a las TICs de los que no. En un mundo interconectado, donde todos los objetos son susceptibles de ser programados y “adaptados”, la nueva “brecha digital” se refiere a quienes tienen competencias en “programación” y los que carecen de nociones básicas de este nuevo segundo lenguaje universal.

En Argentina, existen numerosas iniciativas. Desde Program.AR impulsada por el gobierno nacional hasta otras más pequeñas llevadas a cabo por la motivación e interés de docentes particulares.

En el LINTI, Facultad de Informática, se trabaja en esta línea desde hace 8 años. Luego de lograr afianzar el grupo y generar experiencias replicables, el planteo ahora es la formulación de estrategias y herramientas de evaluación que permita medir el impacto que la enseñanza de la programación tiene en otras áreas de conocimiento en los niños y

jóvenes de escuelas primarias y secundarias.

Palabras clave: enseñanza de programación, robótica educativa, pensamiento computacional.

Contexto

La línea de investigación “Programar en la escuela” presentada en este trabajo, está inserta en el proyecto de investigación “Internet del Futuro: Ciudades Digitales Inclusivas, Innovadoras y Sustentables, IoT, Ciberseguridad, Espacios de Aprendizaje del Futuro” del Programa Nacional de Incentivos a docentes-investigadores, que se desarrolla en el Laboratorio de Investigación en Nuevas Tecnologías Informática (LINTI) de la Facultad de Informática de la Universidad Nacional de La Plata (UNLP). Este proyecto está acreditado por la UNLP y financiado por partidas del presupuesto nacional.

Introducción

Enseñar a programar en la escuela es hoy una preocupación global, algunos países están a la vanguardia del tema, ejemplo de ello es el Reino Unido que desde 2014 ha incorporado la enseñanza de la programación al sistema de educación escolar a partir de los 5 años; Estonia, una de las economías europeas que más apuesta al desarrollo tecnológico, viene impulsando a través de diferentes iniciativas estatales la enseñanza de la programación en la escuela; varios estados de los EEUU también han implementado políticas activas en respuesta al respaldo de la industria tecnológica, a través del movimiento “Learn-To-Code”, liderado fundamentalmente por las iniciativas globales code.org¹ y [codecademy](http://codecademy.com)². En diciembre de 2013 el presidente Barack Obama destacó la visión de este movimiento y alentó a los niños, niñas y jóvenes a participar en el evento anual “La hora del código³” que promueve la educación en “programación” globalmente a través de declaraciones emitidas en un video en el que se enfatiza el hecho de formar jóvenes productores de tecnologías y no meros consumidores: *“No se compren un nuevo videojuego. Hagan uno. No descarguen la última aplicación. Ayuden a diseñar una. No jueguen con el celular. ‘Prográmenlo’”* (Obama, 2013). Israel, Australia, Japón y Finlandia son ejemplos de otros países que cuentan con iniciativas similares en relación de la incorporación de la enseñanza de la informática en las escuelas, en particular sobre contenidos de programación.

1<http://code.org>

2<http://www.codecademy.com>

3<https://code.org/educate/hoc>

Actualmente, el uso de las TIC en las escuelas de Argentina no es algo extraño. Las computadoras llegaron al aula a través de distintos programas de gobierno como el de Programa Nacional Conectar-Igualdad que entregó a la fecha 5.317.247 netbooks a estudiantes y docentes de escuelas secundarias públicas como parte del material escolar. En lo que respecta a educación primaria, un programa similar (Primaria Digital) distribuyen “carritos tecnológicos” con 30 netbooks que quedan en las escuelas y puede ser usado directamente en las aulas. Este tipo de programa está permitiendo superar la primera brecha digital, que es la del acceso a las tecnologías digitales.

También es importante observar que en general las computadoras en las escuelas son utilizadas para navegar, buscar información en Internet y usar aplicaciones de ofimática como procesadores de texto y planillas de cálculo, evidenciando la segunda brecha digital que el padre de la “web”, Tim BernersLee (BernersLee T., 2013) identifica como la que separa a las personas que saben programar de aquellas cuyas habilidades informáticas se limitan simplemente a usar aplicaciones.

Saber programar es un habilitador para entender e intervenir en un mundo cada vez más tecnológico, permitiendo pensar y actuar creativamente ante problemas inesperados (Resnick M., 2008).

Más recientemente, en agosto de 2015 el Consejo Federal de Educación declaró el aprendizaje de programación como una herramienta de “importancia estratégica para el sistema educativo argentino”, que será enseñada durante el ciclo de escolaridad obligatoria en todas las escuelas de la Argentina (Resolución CFE N° 263/15).

Desde el año 2008 el LINTI está trabajando en una línea de investigación cuyo objetivo es promover la enseñanza

de la programación en un principio, en las escuelas secundarias, entendiendo que las habilidades que los estudiantes adquieren programando impactan positivamente en otras áreas disciplinares (Díaz J. et al, 2014). Saber programar contribuye al “pensamiento computacional” (Wing J, 2006), permitiendo a los estudiantes que asuman un papel activo en relación al uso de sus computadoras y de los elementos tecnológicos que los rodean. El “pensamiento computacional” promueve el pensamiento analítico, sistemático, fomenta la creatividad y el trabajo colaborativo, todas ellas habilidades consideradas fundamentales para la sociedad del siglo 21 (Resnick M. et al, 2009). Los estudiantes dejan de ser consumidores de tecnología y pueden dar respuestas imaginativas y divergentes y poner en juego sus significados personales.

En nuestro país la enseñanza de la programación está más difundida en la escuela secundaria, sin embargo su incorporación a edades tempranas ayuda a articular y analizar el conocimiento y a tomar control del proceso de aprendizaje. Seymour Papert, hace más de 30 años decía que programar una computadora no significa ni más ni menos que comunicarse con ella en un lenguaje que tanto la máquina como el usuario 'comprenden'. Y aprender lenguajes es una de las cosas que mejor hacen los niños. Todo niño normal aprende a hablar ¿por qué no aprendería entonces a hablar con la computadora?(Papert, 1980)

Si bien aprender a programar promueve habilidades necesarias en la sociedad moderna, la forma en que se enseña a programar es fundamental. Las teorías de Papert y otros investigadores del área datan de casi medio siglo, y, sin embargo, recién en los últimos años se han vuelto a revisar y han prosperado

numerosas iniciativas al respecto. La forma en que inicialmente se encaró la enseñanza de informática en la escuela hizo que las teorías impulsadas por Papert fueran puestas en acción con cierto retraso.

Líneas de Investigación, Desarrollo e Innovación

A lo largo de los años, se han desarrollado dos propuestas que han convergido en el tiempo. Analizando los resultados de las experiencias llevadas a cabo, se han integrado dos proyectos cuyos objetivos generales siempre fueron similares: promover la enseñanza de la programación en escuelas a través de actividades lúdicas. El proyecto JETS: Java en la Escuela Secundaria que promovía el uso de la herramienta RITA⁴ como una primera aproximación a los conceptos de la programación y luego derivaba en la enseñanza del lenguaje Java, se integró con el proyecto “Programando con Robots y Python”(Díaz J. et al, 2014). el cual plantea actividades directamente a través de un lenguaje de programación real, como lo es Python y el uso de robots físicos inalámbricos.

Combinando los aspectos más destacados de ambos proyectos y los resultados obtenidos de las experiencias en campo, se plantea una nueva línea de investigación que analiza y promueve el desarrollo de herramientas y estrategias pedagógicas que acompañen un único proyecto destinado tanto a educación media como primaria.

Hasta el momento se trabajó con estudiantes de la escuela secundaria solamente y los indicadores formulados están relacionados con la motivación a la

4 <https://github.com/maxirp9/RITA.git>

hora de programar y la elección de carreras en la continuación de sus estudios, pero no se pudo evaluar metódicamente cómo las nuevas habilidades impactan sobre las otras áreas disciplinares y en la propia formación de los estudiantes. Por lo tanto es fundamental generar indicadores y herramientas de evaluación que permitan medir esto último.

“Internet del futuro” abarca el conjunto de servicios y actividades que viabilizan nuevas arquitecturas en el ámbito de “la Red”. El tamaño, la complejidad y el rol que Internet juega en la vida cotidiana tiene un impacto directo en la forma de trabajar, estudiar, aprender, enseñar y relacionarse con otras personas. El desarrollo de aplicaciones inteligentes donde las “cosas” (autos, electrodomésticos, tecnología que se utiliza en salud, en educación, en cuidado del ambiente) interactúan entre ellas y con las personas, propone esta nueva Internet: la de las “cosas”, los servicios y la infraestructura. Pensado en los ciudadanos que formarán parte de esta sociedad del siglo XXI es que se debe trabajar en achicar esta segunda brecha digital propuesta por BernersLee.

Resultados y Objetivos

Las actividades llevadas adelante hasta el momento han arrojado resultados alentadores. En algunas instituciones donde se ha trabajado, los contenidos se han incorporado a la currícula escolar. Ejemplo de esto, son los cursos donde se dicta programación con Python, entre ellos el Colegio Nacional “Rafael Hernández” de la ciudad de La Plata, las escuelas técnicas de EET N° 2 de Ensenada, EET N° 1 y N° 2 de Berisso, la EET N° 6 y N° 8 de La Plata. Lo mismo

ocurre con la EET N° 5 de Berazategui, la EET N° 2 de Berisso y la EEST N° 3 de Mar del Plata, que han incorporado el uso de RITA en algunos de sus cursos.

El trabajo en esta línea dio como resultado también la tesina de Licenciatura en Informática XremoteBot la cual permite la programación remota de los robots físicos del proyecto y RITA en Red que extiende la herramienta RITA original.

Durante el corriente año se trabajará en prácticas profesionalizantes de la tecnicatura en programación con alumnos de 7mo. año de la EEST N° 9 de La Plata. Esto plantea un nuevo desafío, dado que al finalizar las mismas, los estudiantes obtendrán la certificación de técnicos en programación y a lo largo de las prácticas deberán desarrollar un producto de software.

Se comenzará a trabajar en la formulación de indicadores de impacto y el desarrollo de materiales nuevos que se adapten a esta nueva línea.

Con respecto a la inserción de la programación en la escuela primaria, se trabajarán aspectos de robótica educativa y estrategias pedagógicas que combinen la realidad física y la aumentada.

Se planificaron experiencias de campo con niños y niñas de las escuelas primarias Anexa y Lincoln de la ciudad de La Plata.

Formación de Recursos Humanos

El equipo de trabajo que lleva adelante estas líneas forma parte del LINTI.

Así mismo, se han formulado varias tesinas, tesis de postgrado y actividades de cátedras relacionadas con las herramientas involucradas.

Referencias

BernersLee T. (2013) “Saber programación es la nueva brecha digital”, BUSINESS TI | PCWorld España. 05/02/2013.

<http://www.pcworld.es/business-ti/saber-programacion-es-la-nueva-brecha-digital-segun-bernerslee>. [Último acceso 10/03/2016].

Díaz J. et al (2014). Diaz Javier, Banchoff Tzancoff Claudia, Queiruga Claudia, Martín Sofia (2014). Experiencias de la Facultad de Informática en la Enseñanza de Programación en Escuelas con Software Libre. Publicado en Memorias del Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación, Buenos Aires, Argentina, noviembre de 2014. Recuperado de: <http://www.oei.es/congreso2014/memoria/cte/1426.pdf>

Obama (2013). Barack Obama. Discurso difundiendo la hora del código. Computer Science Education Week 2013 Disponible en <https://www.youtube.com/watch?v=6XvmhE1J9PY>

Resnick M. (2008) “Sowing the Seeds for a More Creative Society”. *Learning & Leading with Technology*, 35(4), 18-22.

Resnick M. et al (2009) *Scratch: Programming for All*, Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Amon Millner, Eric Rosenbaum, Jay Silver, Brian Silverman, and Yasmin Kafai.

Communications of the ACM. Vol. 52 No 11, November 2009, 60-67.

Seymour Papert. (1980) *Seymour Papert. Mindstorms: Children, Computers and Powerful Ideas*, Basic Books Co., New York, 1980

Wing J (2006) “Computational thinking”. *Communications of ACM*. Vol 49, N° 3. Mar. 2006, 33–35

Trabajo colaborativo para el diseño, construcción y recuperación de OA accesibles basados en Realidad Aumentada y Robótica pedagógica .

Ana Nieves Rodríguez, Estela Escobar, Leonardo Villatarco, Gisela Soledad Laqihuanaco Ferro, Paola Paniagua

Unidad de Investigación Lide-Ar / Departamento Ciencias de la Computación /
Facultad de Ciencias Exactas / Universidad Nacional de Tucumán
Dirección: Av. Independencia 1800, Teléfonos: 3815305526
e-mails: arodriguez@herrera.unt.edu.ar

Resumen

Palabras clave: Realidad Aumentada, Robótica Pedagógica, Objetos de Aprendizaje

Desde este proyecto, en el que participan la Universidad Nacional de Tucumán y la Universidad Nacional de Colombia, se propone elaborar una conjunto de técnicas que ayuden al docente en la creación de Objetos de Aprendizaje Accesibles, con el agregado de realidad aumentada y la aplicación en Robótica pedagógica. Plantea la combinación de herramientas para lograr la interacción dinámica entre el mundo real y el virtual.

Contexto

Este trabajo se enmarca en el Proyecto de I+D: “Trabajo colaborativo para el diseño, construcción y recuperación de OA accesibles basados en realidad aumentada”.

Se trata de un proyecto en donde participan alumnos, docentes e investigadores de la Universidad Nacional de Tucumán y de la Universidad Nacional de Colombia.

La presencia de los alumnos incorpora la dinámica necesaria para realizar las pruebas con Robots y Realidad Aumentada.

Introducción

Carina Lion [2006] sostiene que “la red Internet ya forma parte de la vida cotidiana, ya sea por la conexión o la imposibilidad de ella, como objeto de luchas y presiones políticas, como discurso pedagógico, etc.” Y plantea preguntas como ¿qué problemáticas trae su incorporación en clase?, ¿cuáles son sus potencialidades?, y con respecto a los estudiantes, ¿sería aconsejable dejarlos navegar y... naufragar?

Otro dato, no menor, es la forma en que los estudiantes se conducen. La simple observación muestra que están continuamente “conectados” mediante sus dispositivos móviles. El alumno es ubicuo, normalmente tiene todo en su celular: calendario, agenda y conexión a internet, desde donde puede leer, recibir mails, noticias de sus amigos en redes sociales, jugar, tomar fotografías y enviar a la red, compartir archivos y trabajar en él en forma colaborativa con compañeros y amigos, escuchar música, grabar una clase y estudiar.

La educación en Argentina se ha insertado en este escenario y los docentes deben cumplir un rol activo en ella. Las políticas que actualmente se están llevando a cabo, muestran la firme

voluntad de incluir a toda la población en el “mundo” de internet. Entre estas políticas se destacan el Programa Conectar-Igualdad, Program.AR, los proyectos de Voluntariado universitario, los proyectos “Universidad Diseño y Desarrollo Productivo”, la Ley de Accesibilidad de la Información en las Páginas Web, sancionada en noviembre de 2010, para promover el acceso a sitios web de las personas con discapacidad. Todas son políticas de inclusión y de atención a la diversidad.

El uso de Internet como fuente de información y comunicación es hoy un complemento indispensable para el sistema educativo.

Los Objetos de Aprendizaje, son entidades digitales con características de diseño instruccional que se encuentran en Repositorios en la Web. Brindan facilidad al adquirir conocimiento de forma casi inmediata y didáctica, buscan acrecentar, en el alumno, la necesidad de continuar investigando sobre alguna temática.

La Realidad Aumentada, como herramienta de apoyo en el ámbito educativo, estimula las ganas de aprender, despierta el interés, aumenta el nivel de atención, crea en los estudiantes un espíritu investigador y muchos otros factores que ayudan a que el entendimiento y asimilación sea mucho más fácil. [Barfield y Caudell]. La Realidad Aumentada incrementa la capacidad innovadora del estudiante, agudiza la atención en el problema planteado, activa la curiosidad y desarrolla la capacidad de investigación de la temática, ya que permite que un modelo interactúe en tiempo y espacio real.

La robótica pedagógica es una disciplina integradora de las distintas áreas del conocimiento, tiene como objetivo instituir las tecnologías en ambientes de aprendizaje mediante la adquisición de habilidades tanto científicas como tecnológicas para la resolución de problemas. Sienta sus bases en el construccionismo de Seymour Papert, que destaca la importancia de la acción, es decir del proceder activo en el proceso de aprendizaje. Postula la necesidad de entregar al alumno herramientas que le permitan construir sus propios procedimientos para resolver una situación problemática. Este aprendizaje otorga importancia al error como detonante de alternativas de solución, permitiéndole transformar las actividades abstractas en concretas, observables, manipulables y controlables.

Sin embargo, aún existen retos que superar:

1- La brecha existente entre el uso y “abuso” de la tecnología por parte de los alumnos y la dificultad o resistencia presentada por parte de algunos docentes.

2- Todos los docentes generan Objetos de Aprendizaje que, frecuentemente, no siguen normas ni son publicados.

3- Los docentes de teoría se quejan, porque a pesar de su esfuerzo, los alumnos se distraen o faltan a sus clases, por lo que se presenta la siguiente hipótesis:

- Es posible despertar el interés de los estudiantes, mediante el uso de Objetos de Aprendizaje con Realidad Aumentada y la incorporación de robótica en la enseñanza?

- La Realidad Aumentada es una técnica de visualización que superpone

información virtual sobre un escenario real; esta superposición se visualiza mediante una pantalla, generalmente de un celular, en donde se mezcla la información en video que capta la cámara con la información virtual creada previamente y es sincronizada a través de marcas o patrones.

○La robótica pedagógica se está destacando como un recurso eficaz para el trabajo interdisciplinario y brinda una mejora en los procesos de enseñanza- aprendizaje. Se caracteriza por el uso o construcción de un pequeño robot que, controlado con software sencillo, permite a los alumnos aprender por ensayo-error a resolver determinadas situaciones problemáticas.

○Los Objetos de Aprendizaje facilitan una educación flexible y personalizada, permitiendo que los estudiantes y docentes puedan adaptar los recursos didácticos de acuerdo con sus necesidades, inquietudes, estilos de aprendizaje y enseñanza.

Líneas de Investigación, Desarrollo e Innovación

- 1.- Diseño de Objetos Virtuales de Aprendizaje
- 2.- Realidad Aumentada
- 3.- Robótica pedagógica

Resultados y Objetivos

Objetivo General

Establecer una metodología para desarrollar Objetos de Aprendizaje accesibles, utilizando tecnologías emergentes como Realidad Aumentada y el uso de la Robótica como herramienta pedagógica.

Objetivos específicos

Estudiar y analizar los distintos enfoques del “Diseño Universal” aplicado a la educación. Relevamiento del estado de la cuestión en la UNT.

Realizar un análisis comparativo de diferentes Metodologías de ingeniería de web para el desarrollo de Objetos de Aprendizaje Accesibles.

Realizar el análisis y definición de estándares de Metadatos y Seleccionar los acordes al proyecto.

Evaluar los requerimientos para un Repositorio de Acceso Abierto y posibilidades de uso del repositorio de la UNT o la UNC

Caracterizar las técnicas de Realidad Aumentada a aplicar en el proyecto

Crear Objetos de Aprendizajes aprovechando tecnologías emergentes, como realidad aumentada

Realizar un estudio minucioso para crear un escenario en donde se combine realidad aumentada y robótica pedagógica, haciendo uso de robots de bajo costo, tipo Lego.

Pruebas realizadas

Las pruebas realizadas en esta primera etapa, se limitan al acercamiento a la programación de robots y a la búsqueda de situaciones problemáticas en las que se pueden aplicar.

Ejemplo1: Robot seguidor de línea negra

Se trabajó con robots con placa arduino, diseñados y ensamblados en Argentina por la empresa Robotgroup.

La programación se realizó en miniBloq V 0.83. Minibloq es un entorno de programación gráfica sencillo. Es fácil de instalar y ha sido especialmente diseñado para quienes están

dando sus primeros pasos en programación, computación física y robótica.

Se trata de robots con capacidad de desplazamiento, acoplados sobre sistemas de propulsión. Se desplazan sobre ruedas u orugas, y son controlados por medio de un mando a distancia, o de forma autónoma utilizando la información captada por sus sensores.

Los robots seguidores de línea se guían por la información recibida de su entorno a través de sus sensores.

Estos Robots se usan para el transporte de piezas de un punto a otro de una cadena de fabricación, suministro de material, transporte de mercadería peligrosa, construcción de túneles, reparaciones de oleoductos, despliegue de cables submarinos, misiones de exploración espacial, submarina o en caso de catástrofes, en algunos casos son usados como artificieros para hacer explotar bombas, realización de tareas domésticas, etc.



Ilustración 1: Programa Robot seguidor de línea



Ilustración 2: Trabajo con sensores

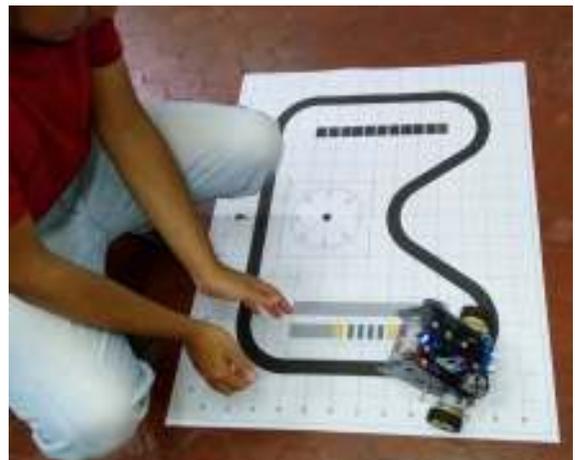


Ilustración 3: Prueba en la pista

Formación de Recursos Humanos

- 1.- Dirección de pasantías estudiantiles y para graduados, dentro del marco de diferentes proyectos de investigación.
- 2.- Dirección de estudiantes en actividades de investigación, presentando y publicando trabajos en jornadas científicas.
- 3.- Dictado de Cursos y Cursos-Talleres Extracurriculares destinados a la

capacitación de docentes a nivel de grado y de posgrado.

3.- Dirección de tesinas de grado

4.- Estudios de posgrado previstos: Maestrías y Doctorados

5.-Trabajo con alumnos en iniciación a la Docencia e Investigación

Referencias

ROSANIGO, ZB. BRAMATI, P. Objetos de aprendizaje: Servicio de Difusión de la Creación Intelectual.SeDiCl. Disponible en Internet:

<<http://sedici.unlp.edu.ar/handle/10915/19934>> . [Consulta: Septiembre de 2015]

Lion, Carina. 2006. "Imaginar con tecnologías. Relaciones entre tecnologías y conocimiento" Editorial La Crujía Ediciones. Buenos Aires, Argentina. 240 páginas. ISBN 987-1004-98-2

López-Carrasco, M.; Tarango, J.; Murguía, P.; Romo, J. R. 2010. "Análisis comparativo del concepto producción científica entre docentes universitarios y organismos evaluadores" Ponencia presentada en las XLI Jornadas Mexicanas de Bibliotecología. Universidad Autónoma de Chihuahua. México.<http://es.scribd.com/doc/37365870/Analisis-comparativo-del-concepto-produccion-cientifica-entre-docentes-universitarios-y-organismos-evaluadores> [Consulta: Agosto de 2015]

D. A. Willey. Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. 2000. Disponible: www.reusability.org/read/chapters/wiley.doc [Consulta: Agosto 2015]

Omar Lucio Cabrera Jiménez. La Robótica Pedagógica. Un vasto campo para la investigación y un nuevo enfoque para la academia. Universidad Tecnológica de Nezahualcóyotl. Disponible: [https://web.archive.org/web/20070320212918/http://fodweb.net/robotica/roboteca/articulos/pdf/](https://web.archive.org/web/20070320212918/http://fodweb.net/robotica/roboteca/articulos/pdf/robotica_pedagogica.pdf)

[robotica_pedagogica.pdf](#) [Consulta: Agosto 2015]

Roberto Garrido, Alex García-Alonso. Técnicas de Interacción para Sistemas de Realidad Aumentada. Disponible: https://www.researchgate.net/publication/228452692_Tcnicas_de_Interaccin_para_Sistemas_de_Realidad_Aumentada [Consulta: Agosto 2015]

Héctor López Pombo. Análisis y Desarrollo de Sistemas de Realidad Aumentada. Disponible: http://eprints.ucm.es/11425/1/memoria_final_03_09_10.pdf [Consulta: Agosto 2015]

TecnoCampus Mataró, UPF. Responsable: Marco Antonio Rodríguez-Xnergie, en colaboración del Departament d'Ensenyament, Generalitat de Catalunya. MOOC "Robots y videojuegos en las aulas: Scratch y Arduino para profesores"

J. Javier Esquivia Mira. Curso Scratch+Arduino. Disponible: <https://www.youtube.com/playlist?list=PL5tA44GpyFntFHtHd6Zxb2xQZpxwOfeu3> [Consulta: Agosto 2015].

BARFIELD & CAUDELL.

T. Fundamentos de Informática usable y Realidad Aumentada. Mahwah, NJ: Lawrence Erlbaum. 2001

Promoción del pensamiento computacional para el fomento de vocaciones TIC y mitigación de índices de desgranamiento en carreras de Informática

Gladys Dapozo; Cristina Greiner, Raquel Petris, María Cecilia Espíndola,
Ana María Company

Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura.
Universidad Nacional del Nordeste

{gndapozo, cgreiner, rpetris, mcespindola, amcompany}@exa.unne.edu.ar

Resumen

Este proyecto de investigación busca promover y difundir el uso de herramientas de programación como forma de introducir aproximaciones al pensamiento computacional en el ámbito de las escuelas secundarias y al inicio de la carrera de Sistemas de la UNNE. Se espera que estas estrategias favorezcan el incremento de jóvenes interesados en seguir carreras de Informática, y dentro de estos, ampliar la participación de las mujeres. Por otra parte, al interior de la carrera, brindar tempranamente a los alumnos herramientas lúdicas que estimulen el pensamiento computacional podría contribuir a mejorar su motivación y comprensión de las técnicas de programación, y de este modo, contribuir a mitigar los preocupantes índices de desgranamiento y de abandono del primer año de carrera.

Palabras clave: Enseñanza de la programación. Vocaciones TIC. Pensamiento computacional.

Contexto

Las líneas de I/D corresponden a un proyecto nuevo presentado a la Secretaría de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE), en etapa de evaluación.

Introducción

En la Argentina, la Fundación Sadosky¹ trabaja en la articulación entre el sistema científico-tecnológico y la estructura productiva en el ámbito de la informática y las telecomunicaciones, generando con ello un impacto positivo en la sociedad y en las posibilidades de desarrollo del país. Uno de sus objetivos es incorporar el estudio de programación en las escuelas argentinas y fomentar el incremento de la matrícula en carreras relacionadas con las Tecnologías de la Información y la Comunicación (TIC). Para ello lleva adelante diversos programas:

El programa Vocaciones en TIC tiene como objetivo principal despertar interés en los jóvenes para estudiar carreras vinculadas con las TIC, en forma más amena y desestructurada, mediante la programación de juegos y animaciones. Como objetivos particulares se definieron:

- Acercar a los alumnos del nivel medio, experiencias prácticas acerca de las actividades propias del quehacer del profesional Informático.
- Incorporar las nociones básicas de programación mediante actividades lúdicas que permiten apoyar la estructura de conocimientos de la disciplina Informática.

Para cumplir estos objetivos se realizan visitas a las escuelas del nivel medio para

¹ <http://www.fundacionsadosky.org.ar/>

realizar talleres de programación, basado en la herramienta Alice/Rebeca, orientados a la elaboración de juegos y animaciones, de manera sencilla y amigable.

La iniciativa Program.AR² tiene como objetivo llevar la enseñanza y el aprendizaje de las Ciencias de la Computación a la escuela argentina. Esta iniciativa incluye múltiples aspectos relacionados con la difusión y popularización de la disciplina, la generación de contenidos escolares y la formación docente, entre otros. La iniciativa es desarrollada de manera conjunta por la Fundación Sadosky, el portal Educ.ar y el Programa Conectar Igualdad.

Por otra parte, la Subsecretaría de Gestión y Coordinación de Políticas Universitarias promueve políticas y acciones para impulsar la integración y articulación del sistema educativo y fortalecer el trabajo conjunto de los niveles secundario y universitario.

El principal objetivo es facilitar el tránsito de los alumnos entre los niveles medio y universitario y evitar la deserción en el primer año de las carreras en la Universidad Pública mediante un trabajo pedagógico internivel e interdisciplinario que permita desarrollar una visión integral de la formación y promover el desarrollo de competencias transversales que faciliten a los alumnos el traspaso entre niveles educativos.

Con este propósito implementaron el Proyecto de Mejora de la Formación en Ciencias Exactas y Naturales en la Escuela Secundaria, destinado a promover el mejoramiento de la calidad de la enseñanza de las ciencias exactas, ciencias naturales y tecnología en el nivel secundario y generar un impacto positivo en la matrícula de ingreso a la Universidad, especialmente en las carreras consideradas estratégicas para el desarrollo del país y comprendidas en la

convocatoria del proyecto. Se financiaron 39 proyectos de 37 Universidades Nacionales y Provinciales, que abordan fundamentalmente 3 ejes de acción: el aseguramiento de las competencias transversales de egreso del nivel secundario, la generación de vocaciones tempranas en el área de ciencias exactas y naturales, y el acompañamiento pedagógico a los docentes para la mejora en la enseñanza en estas áreas. Para llevar adelante este proyecto las Universidades firmaron un acuerdo con la Dirección de Nivel Secundario de cada Jurisdicción para garantizar un trabajo conjunto y articulado [1].

Problemática que enmarca el proyecto

Las carreras de Informática del país, en general, han disminuido el número de ingresantes. Se estima que menos del 5% de los estudiantes eligen carreras de Informática [2]. Esta situación afecta la satisfacción de la demanda de profesionales formados para el crecimiento de la industria.

Además de la disminución de ingresantes se observa también una menor participación de las mujeres. De acuerdo a estudios recientes las mujeres representan sólo un 18% de los estudiantes de informática [3].

Sanz [4] señala que en su evolución la Informática se fue volviendo una ciencia cada vez más matematizada y cercana a las ingenierías o “ciencias duras” lo que significa, en términos de género, más masculinizada. La ausencia femenina en este campo limita la capacidad creativa y la competitividad global, advierten los especialistas, por lo que se encuentran en marcha distintas iniciativas para incorporar mujeres a la Informática [5].

Los estereotipos de género y la falta de modelos femeninos en este campo son factores influyentes, por lo que diversos estudios apuntan a la necesidad de motivar a las chicas en edad temprana [6].

Por otra parte, desde distintas asociaciones académicas, se promueve que la

² <http://program.ar>

Informática se convierta en asignatura troncal en la Educación Secundaria, con estructura y contenidos concretos y orientada al pensamiento computacional, de tal forma que dote a los alumnos de los fundamentos necesarios para comprender qué son y cómo funcionan las computadoras y los programas y aplicaciones que hacen que el mundo de hoy en día funcione. Principalmente, la programación de computadoras como una herramienta para mejorar las habilidades de los alumnos en cuanto a la creatividad y a la resolución de problemas, necesarias en las otras disciplinas [7].

Simari [8] sostiene que la interacción de la Informática con todas las áreas del conocimiento, establece la necesidad de introducir los conocimientos computacionales fundamentales en la formación básica profesional. La realidad del mundo moderno se complementa con un espacio virtual del que es imprescindible conocer los principios elementales para poder comprenderlo y desarrollar tareas en él.

Pensamiento Computacional

El objetivo del Pensamiento Computacional (PC) es desarrollar sistemáticamente las habilidades de pensamiento crítico y resolución de problemas con base en los conceptos de la computación. En suma, potenciar el aprovechamiento del poder de cálculo que tienen las computadoras actualmente

Educar a los estudiantes en PC no significa que necesariamente se desempeñarán en el campo de la ciencia de la computación pero, si lo hacen, estarán mucho mejor preparados para sus cursos universitarios desarrollando habilidades necesarias para comprender y desempeñarse en el mundo.

Wing [9] entiende por pensamiento computacional al proceso que permite formular problemas de forma que sus soluciones pueden ser representadas como secuencias de instrucciones y algoritmos.

La programación permite materializar la idea de abstracción, uno de los procesos claves del pensamiento computacional, e incluso dentro de esta actividad se demuestra que tan útil es dominar esta idea. El pensamiento computacional, entonces, también se hace concreto cuando aprendemos a programar [10].

Los párrafos anteriores muestran la importancia de promover y difundir el uso de herramientas de programación, como forma de introducir aproximaciones al pensamiento computacional, válido para cualquier formación, y para que los jóvenes conozcan sobre la disciplina y consideren una formación específica en la misma. Se espera que estas estrategias favorezcan también el incremento de jóvenes interesados en seguir carreras de Informática, y dentro de estos, ampliar la participación de las mujeres. Al interior de las carreras universitarias de informática, brindar tempranamente a los alumnos herramientas lúdicas que estimulen en pensamiento computacional podría contribuir a mejorar el aprendizaje de las técnicas de programación, y de este modo, contribuir a mitigar los preocupantes índices de desgranamiento y de abandono del primer año de carrera, principalmente.

Líneas de investigación y desarrollo

Esta línea de trabajo se propone realizar un aporte de información para la problemática planteada, tomando como fuente de datos la información de alumnos y docentes de las escuelas del nivel medio de la ciudad de Corrientes y los propios alumnos de la carrera Licenciatura en Sistemas de Información en su etapa inicial.

Las actividades previstas son:

- Continuar con el análisis de los datos que se recaban en los talleres de programación con Alice que se realizan en las escuelas secundarias de Corrientes, en el marco del proyecto de Visitas a las Escuelas, de

la Fundación Sadosky, y en el marco del proyecto de Mejora de la Formación en Ciencias Exactas (SPU).

- Analizar si los factores detectados en los ingresantes se relacionan de alguna forma con la permanencia en la carrera, analizando el avance de los alumnos que ingresaron desde el año 2013 y tuvieron acceso a las actividades de promoción de las vocaciones TIC.
- Evaluar el resultado de la incorporación de actividades con Scratch (Método Sadosky), que a modo de aprestamiento inicial, se introdujo en la asignatura Algoritmos y Estructuras de Datos I de la Licenciatura en Sistemas de Información en el ciclo lectivo 2016. Esta es la primera materia en el trayecto de la programación en el plan de estudio actual.
- Comparar los resultados obtenidos en la etapa de aprestamiento (actividades Scratch), con el rendimiento académico obtenido por los alumnos al finalizar el cursado, para detectar si la actividad de aprestamiento influyó en un resultado final favorable.
- Analizar los datos de una encuesta realizada a los docentes del nivel medio del área tecnología que participaron en el curso de capacitación en didáctica de la programación, con el objetivo de detectar cuestiones favorables o no, que permitan contribuir a modificar el enfoque educativo actual centrado en la capacitación en herramientas utilitarias por otras más específica de la Informática, como la programación.
- Analizar el resultado de las investigaciones vinculadas con esta problemática, realizadas en otras regiones del país, a fin de detectar factores comunes o aspectos particulares en la región.

Resultados y Objetivos

Los principales resultados de las actividades desarrolladas en estas líneas son:

- Se analizó la participación de las mujeres en las distintas carreras científico-tecnológicas de la UNNE en relación a las otras ofertas, en particular el caso de la carrera de Informática. En coincidencia con una tendencia mundial, se comprueba que en esta universidad, decrece el interés de los alumnos por estudiar Informática, y este desinterés es más notorio en las mujeres [11].
- A partir de la información relevada en las actividades de visitas a las escuelas, se realizó un estudio acerca del perfil tecnológico de los alumnos del nivel medio, su interés por seguir carreras vinculadas con la Informática y los factores que influyen en la elección. Los resultados en [12], [13].
- Los resultados de la implementación de los cursos de capacitación docentes se presentaron en la Jornada de Intercambio de Experiencias las TIC en la educación universitaria [14].

Formación de recursos humanos

En esta línea de trabajo están involucrados 3 docentes investigadores y 2 docentes auxiliares del área Programación.

Referencias

- [1] Secretaría de Políticas Universitarias. La Universidad y la Escuela secundaria. Proyecto de Mejora de la Formación en Ciencias Exactas y Naturales. <http://portales.educacion.gov.ar/spu/la-universidad-y-la-escuela-secundaria/>
- [2] Anuario de Estadísticas Universitarias Argentina 2011. ISSN 1850-7514. Departamento de Información Universitaria, de la Secretaría de Políticas Universitarias (SPU), del Ministerio de Educación de la Nación.

- [3] Dr. Mariano Zukerfeld y equipo. "Y las mujeres... ¿dónde están?" Primer estudio sobre la baja presencia femenina en Informática. Fundación Dr. Manuel Sadosky. <http://www.fundacionsadosky.org.ar/wp-content/uploads/2014/06/resumen-mujeres-y-computacion-2013.pdf>
- [4] Sanz, V. (2008). Mujeres e Ingeniería Informática: El caso de la facultad de Informática de la UPM. *ARBOR Ciencia, Pensamiento y Cultura*, CLXXXIV 733. Pag. 905-915. ISSN: 0210-1963.
- [5] Tandon, N. (2012). Oportunidades en materia de TIC: un futuro prometedor para una nueva generación de mujeres. ITU (Unión Internacional de Telecomunicaciones). Switzerland. Disponible en: <http://girlsinct.org/sites/default/files/pages/exec.sum-s.pdf>
- [6] Pérez, P. Estimular en las niñas el interés por la informática erradicará la desigualdad. *Revista Electrónica de Ciencia, Tecnología, Sociedad y Cultura*. ISSN 2174-6850. Disponible en: http://www.tendencias21.net/Estimular-en-las-ninas-el-interes-por-la-informatica-erradicara-la-desigualdad_a7641.html
- [7] Conferencia de Directores y Decanos de Ingeniería Informática (CODDII). La CODDII reclama que la Informática sea asignatura troncal en institutos. *Revista Electrónica de Ciencia, Tecnología, Sociedad y Cultura*. ISSN 2174-6850. Disponible en: http://www.tendencias21.net/La-CODDII-reclama-que-la-Informatica-sea-asignatura-troncal-en-institutos_a35736.html
- [8] Simari, G. (2013). Los fundamentos computacionales como parte de las ciencias básicas en las terminales de la disciplina Informática. Obtenido de SEDICI Repositorio Institucional de la UNLP: <http://sedici.unlp.edu.ar/handle/10915/27579>
- [9] Wing, J. Computational thinking. *Communications of the ACM*, 49(3):33-35, 2006.
- [10] Zúñiga, M., Rosas, M., Fernández, J., & Guerrero, R. (2014). El desarrollo del pensamiento computacional para la resolución de problemas en la enseñanza inicial de la programación. Obtenido de SEDICI: <http://sedici.unlp.edu.ar/handle/10915/41352>
- [11] Dapozo, G.; Greiner, C.; Pedrozo Petrazzini, G.; Chiapello, J. "Género y TIC: Un estudio descriptivo en la UNNE". III Jornadas de Investigación en Ingeniería del NEA y países limítrofes: Nuevos escenarios para la ingeniería en el Norte Grande. UTN - Facultad Regional Resistencia. ISBN: 978-950-42-0157-1. 9 y 10 de Junio de 2014. Resistencia, Chaco, Argentina.
- [12] Dapozo, G.; Greiner, C.; Pedrozo Petrazzini, G.; Chiapello, J. "Vocaciones TIC. ¿Qué tienen en común los alumnos del nivel medio que tienen interés por carreras de Informática?". IX Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2014). ISBN 978-987-24611-1-9. Pag. 128 a 137. Universidad Nacional de Chilecito. Chilecito. La Rioja. 12 y 13 de junio de 2014.
- [13] Dapozo, G.; Greiner, C.; Pedrozo Petrazzini, G.; Chiapello, J. "Investigación para fortalecer actividades de promoción y retención de alumnos en carreras de Informática". *Anales del XX Congreso Argentino de Ciencias de la Computación*. 1a ed. San Justo. Buenos Aires. Universidad Nacional de La Matanza, 2014. ISBN 978-987-3806-05-6. 23 de octubre de 2014.
- [14] Dapozo, G., Espíndola, M.E.C., Company, A. M., Vallejos, S., Badaracco, Numa. "Desarrollo del pensamiento computacional con SCRATCH". Jornada de Intercambio de Experiencias las TIC en la educación universitaria. 5 de noviembre de 2015. Campus Resistencia.

Reuso Orientado a Servicios: Selección y Testing de Servicios

Andrés Flores, Alejandra Cechich, Martín Garriga,
Marcelo Moyano, Alan De Renzis, Diego Anabalon
Grupo de Investigación en Ingeniería de Software del Comahue (GIISCo)
<http://giisco.uncoma.edu.ar>
Departamento de Ingeniería de Sistemas – Facultad de Informática
Universidad Nacional del Comahue
Buenos Aires 1400, (8300) Neuquén
Contacto: [andres.flores, alejandra.cechich]@fi.uncoma.edu.ar

Resumen

El reuso de artefactos software brinda oportunidades para proveedores y clientes, tanto para acelerar el proceso de desarrollo de software como para establecer oferta de productos reusables. El paradigma de Computación Orientada a Servicios (SOC), promueve el desarrollo de aplicaciones distribuidas en ambientes heterogéneos, que son construidas ensamblando o componiendo servicios reusables, que se publican a través de una red y se acceden mediante protocolos específicos. SOC ha sido ampliamente adoptado bajo su implementación con la tecnología de Servicios Web, que provee flexibilidad de ejecución remota que oculta las plataformas específicas de ejecución y permite descentralizar los procesos de negocios. SOC requiere la publicación de servicios en un registro (UDDI de acuerdo a Servicios Web), los cuales luego son identificados y evaluados para una aplicación en desarrollo. Sin embargo, aún este proceso necesita métodos exhaustivos y eficientes, tanto para identificación como para selección de servicios, en el cual se puede considerar la aplicación de técnicas de Pruebas de Software y el uso de dos conceptos actuales: Orquestación y Coreografía de servicios.

Palabras Clave: Ingeniería de Software basada en Reuso – Software Orientado a Servicios – Servicios Web – Calidad de Software – Verificación y Validación.

Contexto

La línea presentada se inserta en el contexto de los siguientes proyectos y acuerdos de cooperación:

- 04/F001: “Reuso Orientado a Servicios”, sub-proyecto del Programa “Desarrollo Orientado a Reuso”. Financiado por UNCo. (2013-2016).
- PICT-2012-0045: “Mecanismos y Herramientas para Grid Híbridos Orientados a Servicios”. Financiado por ANPCyT. (09/2013-08/2016).
 - Investigaciones conjuntas con ISISTAN-UNICEN, Tandil.
- Acuerdo de Cooperación con el Grupo Alarcos, Escuela Superior de Informática, Universidad de Castilla-La Mancha, España.

Introducción

Actualmente la industria de software observa cómo mediante el reuso de software se puede alcanzar un proceso de desarrollo de software acelerado y confiable al basarse en artefactos software que ya han sido probados en diferentes contextos de aplicación. Para ello se

adopta el concepto denominado “tercerización”, por medio del cual se acuerdan contratos para adquisición y provisión de artefactos software reusables y se establecen relaciones comerciales entre vendedores y clientes. Por lo tanto, desde el punto de vista de un cliente implica la posibilidad de acelerar el desarrollo de un producto software para reducir el lanzamiento al mercado, y desde el punto de vista de un proveedor implica la posibilidad de observar sus productos con una perspectiva nueva que los coloque dentro del mercado de artefactos reusables.

Un paradigma que promueve altamente el reuso de software se denomina Computación Orientada a Servicios (SOC), donde la funcionalidad a ser reusada adopta la forma de servicios, o unidades lógicas que presentan entornos heterogéneos de ejecución y pueden ser ensambladas para formar otras unidades lógicas de mayor nivel de abstracción que resuelvan (directamente o en parte) los procesos de negocios para un contexto de aplicación [SH05,PTDL07]. El paradigma SOC encontró una plataforma potencial de aprovechamiento mediante la Web, desde donde se desarrolló la tecnología de Servicios Web [NSS03, Wetal05], cuya base es el esquema estándar XML y el Lenguaje de Descripción de Servicios Web (WSDL), que facilitan ampliamente el desarrollo y mantenimiento de especificaciones formales de servicios. Así el paradigma SOC bajo la implementación con Servicios Web ha logrado su amplia adopción en la industria, principalmente bajo la flexibilidad de ejecución remota que permite a las compañías descentralizar aún más sus procesos de negocios y la ventaja de que las plataformas específicas de ejecución se encuentran ocultas, por lo cual no se requiere de inversiones adicionales en tecnología (incluyendo

costos y esfuerzo de aprendizaje) al adquirir funcionalidad de terceras partes. El beneficio que la tecnología de Servicios Web provee al paradigma SOC se ha denominado “relación sin responsabilidad”, donde una aplicación cliente no requiere asumir cómo se ha implementado el servicio con el que se comunica. Sin embargo, los proveedores de servicios tienen la responsabilidad de evaluar la calidad de los productos ofrecidos como servicios y los consumidores de servicios a su vez deben ser capaces de identificar tal calidad que influirá sobre las aplicaciones en desarrollo. Para ello, las estrategias de Pruebas de Software deben ajustarse a este contexto específico donde los servicios están acordados como cajas negras que sólo permite evaluar el comportamiento y cualidades observables externamente [BDN10, Z08].

El funcionamiento concreto del paradigma SOC se basa en la Arquitectura orientada a Servicios (SOA) [SH05] que se encuentra compuesta por tres actores principales: un proveedor, un consumidor y un registro de servicios; donde el proveedor desarrolla y publica servicios en el registro, para que luego el consumidor busque servicios y establezca una comunicación con el proveedor. Sin embargo, la búsqueda de servicios publicados en un registro UDDI (según la tecnología de servicios Web) [OASIS04], en general requiere invertir un esfuerzo considerable para distinguir servicios candidatos que satisfagan los requerimientos de la aplicación cliente [NSS03, Wetal05]. En particular, cuando varios candidatos ofrecen funcionalidades similares se requieren métodos eficientes de selección de servicios que discriminen tanto aspectos funcionales como no-funcionales, considerando además las interacciones válidas para un servicio candidato en función de los procesos de

negocio que implementará la aplicación cliente. En particular el ensamblaje de servicios considerando procesos de negocio e interoperabilidad de servicios plantea el uso de dos conceptos de reciente investigación: Orquestación y Coreografía de servicios [P03, Wetal05]. El primero relacionado a una aplicación particular que describe un proceso de negocios específico, y el segundo relacionado a las interacciones válidas que pueden ocurrir entre distintos servicios predestinados a intervenir en una colaboración. En este contexto se cuenta actualmente con diversos lenguajes de descripción y frameworks de ejecución, tales como Microsoft XLANG (para BizTalk Server) [Micro10], IBM WSFL (Web Services Flow Language) [Snell10], de lo cual surgió el estándar BPEL4WS [OASIS07], y otro estándar WSCDL (Web Services Choreography Description Language) [W3C05] promovido por Sun, SAP, BEA e Intalio.

Líneas de Investigación, Desarrollo e Innovación

El perfil de esta línea puede definirse en base a las actividades de investigación y transferencia, a las que da soporte el grupo GIISCo. Los temas específicos consideran los desafíos diferentes asociados al crecimiento de la Tecnología de la Información y las Comunicaciones. Actualmente, abordamos los siguientes aspectos:

- Evaluación y selección de servicios.
- Adaptación y Composición de servicios.
- Testing de servicios.
- Herramientas para evaluación, selección, composición y testing de servicios.

- Definición de aplicaciones en dominios específicos.

Resultados y Objetivos

En [FCGMR15] hemos enumerado una serie de contribuciones anteriores. Durante el año 2015, hemos profundizado la investigación en aspectos de *evaluación y selección de servicios*, generando métodos y herramientas enfocados en compatibilidad de interfaces y de comportamiento en base a testing. Este avance se ha efectuado en colaboración con investigadores del Grupo Alarcos y de ISISTAN (UNICEN) [AdRBCF15, AGFCZ15a, AGFCZ15b, AGFCZ15c, dRGFCZ15, FCGMR15, GFCZ15, LGdRFMZ15].

Las líneas de investigación convergen en el tratamiento del desarrollo de software basado en el reuso de servicios desde la perspectiva de las aplicaciones orientadas a servicios. Una aplicación orientada a servicios implica una solución de negocio que consume servicios de uno o más proveedores y los integra en un proceso de negocio [SW04]. Además puede verse como una aplicación basada en componentes que integra dos tipos de componentes: internos localmente empotrados en la aplicación, y externos estática o dinámicamente enlazados a algún servicio [CMZC14]. No solamente se enfocará en el reuso de servicios individuales, sino también en la composición de servicios como forma de tercerizar una funcionalidad. Se adoptará la visión de proceso de negocio para la definición de comportamiento, donde se aplicará testing de servicios para una evaluación dinámica. Se complementará el modelo de selección y composición de servicios mediante las últimas plataformas y avances tecnológicos incluyendo semántica y estandarización.

Se prevee la aplicación de estos modelos y las herramientas de soporte a dominios específicos, con particular énfasis en aquellos que requieran rigurosidad como aporte de validación efectiva. La visión de esta línea de investigación se resume en:

“Definir técnicas y herramientas para la mejora del desarrollo de software, en función del reuso de servicios web. La definición de modelos de identificación y selección de servicios, y la posibilidad de composición de servicios”.

Formación de Recursos Humanos

Este proyecto se compone de 9 investigadores, entre los que se cuentan docentes y alumnos del Grupo GIISCO de UNComa y asesores externos. Algunos de los docentes-investigadores se encuentran realizando carreras de postgrado. Se cuenta actualmente con 3 doctores (1 de ellos finalizó a comienzos de 2016, y 1 investigador asistente CONICET), 1 postdoctorando (recién recibido, becario CONICET), 2 doctorandos (becarios CONICET) y 1 maestrando entre los miembros del proyecto. Dirección de Tesis de Grado durante 2014: 3 tesis (1 finalizada a fines de 2015), y se dirigió 1 alumno becario PNB-TICs.

Referencias

- [AdRBCF15] Arias, M., De Renzis, A., Buccella, A., Cechich, A., Flores, A. (2015). *Búsqueda de Servicios para Asistir en el Desarrollo de una Línea de Productos de Software*. ASSE'15, 16th Argentine Symposium on Software Engineering, part of 44 JAIIO, Argentine Conference of Informatics. SADIO. Buenos Aires. pp 145-158.
- [AGFCZ15a] Anabalon, D., Garriga, M., Flores, A., Cechich, A., Zunino, A. (2015). *Adaptability-based Service Behavioral Assessment*. CACIC'15, XXI Congreso Argentino de Ciencias de la Computación. Red-UNCI. Junín, Buenos Aires. p 10.
- [AGFCZ15b] Anabalon, D., Garriga, M., Flores, A., Cechich, A., Zunino, A. (2015). *Test Reduction for Easing Web Service Integration*. ASSE'15, 16th Argentine Symposium on Software Engineering, part of 44 JAIIO, Argentine Conference of Informatics. SADIO. Buenos Aires. pp 115-129.
- [AGFCZ15c] Anabalon, D., Garriga, M., Flores, A., Cechich, A., Zunino, A. (2015). *Adaptability-based Service Behavioral Assessment*. Journal of Computer Science & Technology, vol 15, no 02, pp 75-80. ISTECS. [Latindex].
- [BDN10] Baresi, L.; Di Nitto, E. (2010). *Test and Analysis of Web Services*. Springer.
- [CMZC14] Crasso, M., Mateos, C., Zunino, A., Campo, M. (2014). *EasySOC: Making Web Service Outsourcing Easier*. International Journal on Information Sciences, vol. 259, pp. 452-473.
- [dRGFCZ15] De Renzis, A., Garriga, M., Flores, A., Cechich, A., Zunino, A. (2015). *Case-based Reasoning for Web Service Discovery and Selection*. CLEI'15, XLI Conferencia Latinoamericana en Informática. IEEE Xplorer Digital Library. pp 18-29. Arequipa, Perú. Octubre.
- [FCGMR15] Flores, A.; Cechich, A.; Garriga, M.; Moyano, M.; De Renzis, A. (2015). *Reuso Orientado a Servicios: Mejoras en Evaluación y Testing*. WICC'15, XVII Workshop de Investigadores en Ciencias de la Computación, Salta, Argentina.

- [GFCZ15] Garriga, M., Flores, A., Cechich, A., Zunino, A. (2015). *Web Services Composition Mechanisms: A Review*. IETE Technical Review, ahead-of-print, pp 1-8. Taylor & Francis. (Indexed SCI, IF JCR2013: 0,925).
- [LGdRFMZ15] Lizarralde, I., Garriga, M., De Renzis, A., Flores, A., Mateos, C., Zunino, A. (2015). *Un Enfoque Semántico-Estructural para Selección de Servicios Web*. CIAWI'15, 13th Conferencia Ibero-Americana WWW / Internet 2015. IADIS. Florianópolis, Brasil. Noviembre.
- [Micro10] Microsoft Corporation (2010). *XLANG/s Language*. MSDN Library. [http://msdn.microsoft.com/en-us/library/aa577463\(BTS.70\).aspx](http://msdn.microsoft.com/en-us/library/aa577463(BTS.70).aspx)
- [NSS03] Nagappan, R.; Skoczylas, R.; Sriganesh, R. (2003). *Developing Java™ Web Services: Architecting and Developing Secure Web Services Using Java*. Wiley Publishing Inc.
- [OASIS04] OASIS Consortium (2004). *UDDI Version 3.0.2*. UDDI Spec Technical Committee Draft, October.
- [OASIS07] OASIS Standard (2007). *Web Services Business Process Execution Language Version 2.0*. <http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html>
- [P03] Peltz, C. (2003). *Web Services Orchestration and Choreography*. IEEE Computer, 36(10):46–52.
- [PTDL07] Papazoglou, M.; Traverso, P.; Dustdar, S.; Leymann, F. (2007). *Service-Oriented Computing: State of the Art and Research Challenges*. IEEE Computer 40(11): 38-45.
- [SH05] Singh M.; Huhns, M. (2005). *Service-oriented computing: Key concepts and principles*. IEEE Internet Computing, 9(1):75–81.
- [Snell10] Snell, J. (2010) *The Web services insider, Part 4: Introducing the Web Services Flow Language*. IBM Emerging Technologies. <http://www.ibm.com/developerworks/library/ws-ref4/index.html>
- [SW04] Sprott, D.; Wilkes, L. (2004). *Understanding Service-Oriented Architecture*. The Architecture Journal. MSDN Library. Microsoft Corporation, <http://msdn.microsoft.com/en-us/library/aa480021.aspx>
- [W3C05] W3C Candidate Recommendation (2005) *Web Services Choreography Description Language Version 1.0*. <http://www.w3.org/TR/ws-cdl-10/>
- [Wetal05] Weerawarana, S.; Curbera, F.; Leymann, F.; Storey, T.; Ferguson, D. (2005). *Web Services Platform Architecture: SOAP, WSDL, WS-Policy, WS-Addressing, WS-BPEL, WS-Reliable Messaging, and More*. Prentice Hall PTR.
- [Z08] Zhou, X. (2008). *Testing and Verifying Web Services. From the Researcher's Perspective*. VDM Verlag.

El modelo de negocio en AOP4ST

Fernando Pincioli, Laura Zeligueta

Instituto de Investigaciones
Facultad de Informática y Diseño
Universidad Champagnat
Mendoza, Argentina, +54 (261) 424-8443
{pincirolifernando, zeliguetalaura}@uch.edu.ar

Resumen

Los diferentes paradigmas de desarrollo de software por lo general nacen a partir de los lenguajes de programación y de las tecnologías que permiten aplicar sus ideas. La definición de estos paradigmas continúa siguiendo las fases del ciclo de vida en sentido aguas arriba. El paradigma de orientación a aspectos no es ajeno a esta forma de evolución y desarrollo, por lo que en la fase de modelado de negocio todavía no existe una cantidad de propuestas lo suficientemente numerosas y con gran consenso [1]. Adicionalmente, de las propuestas existentes, solo algunas de ellas emplean el estándar *BPMN* en forma pura, de modo que su aplicación en la industria no es sencilla por las obvias razones de ausencia de herramientas de soporte informático para las actividades de modelado, monitoreo, control y mejora de procesos, etc. necesarias [2].

El objetivo de nuestro proyecto de investigación consiste en establecer un conjunto de reglas que permitan el empleo del paradigma orientado a aspectos en la fase de modelado de negocio mediante el empleo de *BPMN* en forma totalmente estándar.

Esto facilitaría una transición suave desde las prácticas de modelado de negocio actuales hacia las necesarias para aplicar el paradigma orientado a aspectos.

Palabras clave: orientación a aspectos, procesos de desarrollo de software, early aspects, modelado de negocio, *BPMN*, *AO4BPMN*, separación de incumbencias, composición de incumbencias, resolución de conflictos, *AOP4ST*.

Contexto

El Instituto de Investigaciones de la Facultad de Informática y Diseño de la Universidad Champagnat centra sus trabajos de investigación principalmente en el área de Ingeniería de Software, enfocándose en la línea de investigación de “*Desarrollo de Software Orientado a Aspectos*”. Esta línea se inició con el proyecto de investigación “*Procesos de desarrollo de software de calidad basados en aspectos*”, llevado a cabo en la UTN Facultad Regional Mendoza, y que luego continuó en nuestra Universidad con el proyecto “*Definición de criterios para la detección temprana de aspectos en el modelado de negocios y el desarrollo de los requisitos*” que culminó exitosamente el año pasado y fue presentado en WICC 2015. Nuestro proyecto actual recibe aportes de la Universidad Champagnat y de la empresa Aconcagua Software Factory S.A. de la provincia de Mendoza.

Introducción

Los beneficios del paradigma orientado a aspectos pueden ser aprovechados en todas las fases del ciclo de vida del desarrollo de software. Cada una de estas fases tiene objetivos puntuales y específicos, pero tenemos la certeza de que con un enfoque orientado a aspectos sus resultados se pueden optimizar y potenciar. La causa principal que aporta el paradigma para lograr estos beneficios es la *separación de incumbencias* a lo largo de todo el ciclo de vida.

La separación de incumbencias nació para detectar y encapsular aspectos que están *desparramados y enredados* en el código, lográndose de este modo sistemas más modulares, mantenibles, extensibles, reusables, comprensibles, etc. En el modelo de negocio también es posible lograr esos mismos objetivos por medio de la separación de las cuestiones que son específicas del dominio del problema de aquellas otras que no lo son, razón por las que se las conoce como *incumbencias transversales*, por estar presentes no solamente en la mayoría de los procesos de negocio a los que atraviesan en forma transversal, sino también a través de diferentes dominios de problema.

Así, tras detectar las *incumbencias transversales* en el modelo de negocio es posible encapsularlas para separar el conocimiento específico del dominio del problema de aquellos otros elementos que no pertenecen a él.

Esto que acabamos de mencionar se conoce como *enfoque asimétrico*, ya que se hace una diferencia entre los intereses específicos del dominio del problema de aquellos que no lo son, pero existe también un *enfoque simétrico*, que consiste en la separación de las incumbencias que pertenecen al propio dominio de problema. Las incumbencias del primer enfoque se

encapsulan como *aspectos impares* y las del segundo como *aspectos pares*.

Si bien se puede observar una aparente inclinación hacia el enfoque asimétrico, las prácticas habituales de modelado de procesos de negocio también van preparando el terreno para la detección y el encapsulamiento de los aspectos pares.

Además de la detección y encapsulamiento de las incumbencias transversales, tendremos en cuenta su posterior composición en base a reglas, la resolución de los posibles conflictos y el posicionamiento de nuestra propuesta en el concierto de las propuestas existentes.

Líneas de Investigación, Desarrollo e Innovación

Los proyectos de investigación que se vienen sucediendo en nuestra Facultad presentan cuatro ejes de investigación, donde los tres primeros se ven integrados con el cuarto:

1. *Modelado de procesos de negocio orientados a aspectos.*
2. *Separación y composición de incumbencias con resolución de conflictos.*
3. *Especificación de requisitos y gestión de incumbencias con casos de uso.*
4. *Procesos de desarrollo de software orientados a aspectos.*

Resultados y Objetivos

Nuestro proyecto de investigación anterior, centrado en los procesos de desarrollo de software orientados a aspectos, estableció el marco conceptual y evidenció cuestiones adicionales a resolver. De esta manera, en este nuevo proyecto decidimos profundizar en la etapa de *modelado de procesos de negocio* con la finalidad de asegurar que la orientación a aspectos se puede emplear como un

medio eficaz –no como un fin–, para potenciar los objetivos y resultados de esa etapa tan importante del ciclo de vida del desarrollo de software y la primera capa para la construcción de una *arquitectura empresarial*.

Partimos de los trabajos de Jalali [1] [3] [4] [5] [6] [7] con respecto al *modelado de procesos de negocio orientado a aspectos*, quien utiliza la notación *AO4BPMN* propuesta por Charfi et al. [8], hoy ampliamente aceptada y que nosotros consideramos muy adecuada, pero a la que tuvimos que realizar algunos ajustes para que se mantenga dentro de la notación estándar *BPMN 2.0* y para poder indicar la inserción de los advices en múltiples join points en forma simultánea al estilo *AspectJ*.

Para la *detección de incumbencias* en esta etapa de modelado de procesos de negocio estamos analizando diversas técnicas de *minería de aspectos* que, si bien la mayoría de ellas fueron diseñadas para la *refactorización de aspectos* en el código y solo algunas se emplean en las etapas tempranas (*early aspects*), las estamos analizando y adaptando para su empleo [3] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19]. Nos fue aceptado un artículo sobre esta temática para la edición de mayo de este año de la revista “Perspectivas em Ciências Tecnológicas - Revista do Curso de Ciência da Computação” [20].

Es importante destacar que la detección de incumbencias impares es más sencilla, pero también estamos considerando la detección de incumbencias pares, de modo de poder aplicar los *enfoques asimétrico y simétrico* desde esta etapa del ciclo de vida [21] [22].

Tras la separación de incumbencias y su refactorización, hemos desarrollado ideas que nos permitan alcanzar dos objetivos: analizar los modelos completamente

integrados y detectar y resolver los posibles conflictos que podrían aparecer.

Con respecto al análisis de los modelos integrados, estamos elaborando un conjunto de reglas de composición. Para ello estamos recorriendo la totalidad de la notación de BPMN 2.0, incluyendo la combinación de elementos, la vista dinámica de los procesos, los procesos con instancias múltiples y los patrones de modelado de procesos de negocio, sumado a la revisión de una amplia casuística que nos permita emitir nuestras opiniones adecuadamente.

En cuanto a la resolución de conflictos, hemos tomado como punto de partida las propuestas de diversos autores sobre resolución de conflictos en la orientación a aspectos [15] [23] [24] [25], más la resolución de conflictos que ya se da por naturaleza entre los requerimientos [26] [27] [28] [29], pero esto será enriquecido tras los resultados que obtengamos con el trabajo en las reglas de composición mencionadas en el punto anterior. Nos fue aceptado un trabajo al respecto que será publicado por Elsevier Science en Procedia Computer Science series on-line [30], dentro de la 7th International Conference on Ambient Systems, Networks and Technologies, a realizarse del 23 al 26 de mayo próximo en Madrid.

Para el análisis de las relaciones entre aspectos hemos avanzado con dos modelos específicos. El primero de ellos es el *modelo de join points*, que permite analizar los puntos de los procesos en los que se insertan los advices, la relación entre advices que se componen y la posibilidad de alertar sobre los puntos de los procesos en los que se deben aplicar soluciones ante conflictos que fueran aplicadas en otros puntos de los procesos. El segundo modelo es el *modelo de incumbencias*, que presenta la relación entre incumbencias al nivel de los procesos, también para favorecer el análisis y toma de decisiones

para la mejora de las incumbencias detectadas hasta ese momento.

Por último, hemos tomado la comparación de técnicas de modelado de procesos de negocio orientadas a aspectos planteada por Jalali [31] que nos sirve de base para evaluar nuestra propuesta, asegurar que cuente con los criterios más importantes que debería tener y a dotarla de aquellas otras características que nos permitan ofrecer una propuesta superadora con respecto a aquellas evaluadas.

Es importante destacar que nuestro proyecto incluye la aplicación práctica de sus ideas y técnicas en desarrollos de software reales de complejidad diversa, para empresas y organizaciones de envergadura, de una empresa de ingeniería de software de la ciudad de Mendoza y también que continúa con las líneas de investigación establecidas en la Facultad de Informática y Diseño de nuestra Universidad Champagnat.

Formación de Recursos Humanos

El proyecto apunta a la formación del equipo de profesores del área de Ingeniería de Software de la Universidad Champagnat y de los alumnos y egresados participantes tanto del proyecto en sí como de sus avances y resultados.

AOP4ST es el tema central de la tesis del doctorando Fernando Pincirolí, realizando el Doctorado en Ciencias Informáticas de la Universidad Nacional de San Juan, bajo la dirección del Dr. Raymundo Forradellas, de la Universidad Nacional de Cuyo, y la codirección del Dr. José Luis Barros Justo, de la Universidad de Vigo, España.

También se están elaborando la tesis de Gustavo Albino, de la Maestría en Ingeniería de Software de la Universidad Nacional de San Luis, y dos tesinas de grado de alumnos de nuestra Universidad.

Referencias

- [1] A. Jalali, P. Wohed, C. Ouyang, and P. Johannesson, "Dynamic Weaving in Aspect Oriented Business Process Management," *Move to Meaningful Internet Syst. OTM 2013 Conf.*, vol. 8185, pp. 2–20, 2013.
- [2] B. Johansson, B. Andersson, and N. Holmberg, *Perspectives in Business Informatics Research*. Springer, 2014.
- [3] A. Jalali, "Aspect Mining in Business Process Management," *Lect. Notes Bus. Inf. Process.*, vol. 194, pp. 246–260, 2014.
- [4] A. Jalali, "Assessing Aspect Oriented Approaches in Business Process Management," *Lect. Notes Bus. Inf. Process.*, vol. 194, pp. 231–245, 2014.
- [5] A. Jalali, "Foundation of Aspect Oriented Business Process Management," p. 98, 2011.
- [6] A. Jalali, "Enactment, Supporting Oriented, Aspect Process, Business," 2013.
- [7] A. Jalali, P. Wohed, and C. Ouyang, "Aspect Oriented Business Process Modelling with Precedence," 2012.
- [8] A. Charfi, H. Müller, and M. Mezini, "Aspect-oriented business process modeling with AO4BPMN," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6138 LNCS, pp. 48–61, 2010.
- [9] A. Sampaio, N. Loughran, A. Rashid, and P. Rayson, "Mining Aspects in Requirements," *Asp. Requir. Eng. Archit. Des. Work. (held with AOSD 2005), Chicago, Illinois, USA, 2005*.
- [10] S. Clarke and E. Baniassad, *Aspect-oriented analysis and design. The Theme approach*. Boston: Addison-Wesley, 2005.
- [11] A. Rashid, P. Sawyer, A. Moreira, and J. Araújo, "Early aspects: a model for aspect-oriented requirements

- engineering,” *Proc. IEEE Jt. Int. Conf. Requir. Eng.*, pp. 0–3, 2002.
- [12] A. Rago, E. S. Abait, C. Marcos, and A. Díaz-Pace, “Early aspect identification from use cases using NLP and WSD techniques,” *15th Work. Early Asp. - EA '09*, pp. 19–26, 2009.
- [13] A. Rago and C. Marcos, “Uncovering Quality-attribute Concerns in Use-case Specifications via Early Aspect Mining.”
- [14] A. Sampaio, P. Rayson, A. Rashid, and R. Chitchyan, “EA-Miner : a Tool for Automating Aspect-Oriented Requirements Identification,” *Proc. 20th IEEE/ACM Int. Conf. Autom. Softw. Eng.*, pp. 352–355, 2005.
- [15] A. Sardinha, R. Chitchyan, J. Araújo, A. Moreira, and A. Rashid, “Conflict Identification with EA-Analyzer,” in *Aspect-Oriented Requirements Engineering*, 2013, pp. 209–224.
- [16] G. Cojocar, “Aspect mining. past, present, future,” vol. LVII, no. 4, pp. 85–96, 2012.
- [17] R. R. McFadden and F. Mitropoulos, “Aspect mining using model-based clustering,” *2012 Proc. IEEE Southeastcon*, no. 978, pp. 1–8, 2012.
- [18] J. Huang, Y. Lu, and J. Yang, “Aspect Mining Using Link Analysis,” *2010 Fifth Int. Conf. Front. Comput. Sci. Technol.*, pp. 312–317, 2010.
- [19] K. Mens, A. Kellens, and J. Krinke, “Pitfalls in aspect mining,” *Proc. - Work. Conf. Reverse Eng. WCRE*, pp. 113–122, 2008.
- [20] F. Pincioli, “Considerações Acerca da Mineração de Aspectos,” *Perspectivas em Ciências Tecnológicas*, vol. 5, 2016.
- [21] J. Bálik and V. Vranić, “Symmetric aspect-orientation: some practical consequences,” *Proc. 2012 Work. Next ...*, pp. 7–11, 2012.
- [22] D. C. Collell, “Aspect-Oriented Modeling of Business Processes,” no. September, 2012.
- [23] V. Vanoli and C. Marcos, “Early Conflicts : Análisis y Resolución de Conflictos Tempranos.”
- [24] A. Sardinha and A. Moreira, “Conflict Management in Aspect-Oriented,” 2010.
- [25] J. Pryor and C. Marcos, “Solving Conflicts in Aspect-Oriented Applications,” *Framework*.
- [26] K. Wiegers and J. Beatty, *Software Requirements*. 2013.
- [27] L. Chung, B. Nixon, and E. Yu, “Using Non-Functional Requirements to Systematically Select Among Alternatives in Architectural Design,” *1st Int. Work. Archit. Softw. Syst. - Coop. with 17th Int. Conf. Softw. Eng. ICSE 1995*, pp. 31–43, 1995.
- [28] L. Chung, B. Nixon, E. Yu, and J. Mylopoulos, “The NFR Framework in Action,” *Non-Functional Requir. ...*, pp. 15–45, 2000.
- [29] D. Mairiza, D. Zowghi, and N. Nurmuliani, “Managing conflicts among non-functional requirements,” pp. 11–19, 2009.
- [30] F. Pincioli, “Improving software applications quality by considering the contribution relationship among quality attributes,” *Procedia Comput. Sci. 3rd Int. Work. Comput. Antifragility Antifragile Eng. (ANTIFRAGILE 2016)*, 2016.
- [31] L. Lace, R. Liepiņš, and E. Rencis, *Perspectives in Business Informatics Research*, vol. 128, no. September. 2012.

Reuso Orientado a Dominios: Hacia un Proceso Integral de Desarrollo de Líneas de Productos de Software

Agustina Buccella, Juan Luzuriaga, Alejandra Cechich,
Rodolfo Martínez, Rafaela Mazalu, Marcos Cruz,
Matias Pol'la, Maximiliano Arias y Adriana Martin
GIISCO Research Group
Departamento de Ciencias de la Computación
Universidad Nacional del Comahue
Neuquen, Argentina
agustina.buccella@fi.uncoma.edu.ar

Maria del Socorro Doldan y Enrique Morsan
Instituto de Biología Marina y Pesquera "Almirante Storni"
Universidad Nacional del Comahue
Ministerio de Producción de Rio Negro
San Antonio Oeste, Argentina
{msdoldan, qmorsan}@gmail.com

Tamara Rubilar
Centro Nacional Patagónico - CENPAT-CONICET
Puerto Madryn, Chubut, Argentina
tamararubilar@gmail.com

1. Resumen

El reuso de software en general abarca todas aquellas técnicas, procesos y metodologías que tienen por objetivo la reutilización de artefactos de software, creados en cualquiera de las etapas del ciclo de vida, para ser utilizados en nuevos desarrollos o en la construcción de nuevas versiones. Los beneficios a los cuales se debe arribar, a grandes rasgos, son la mejora de los tiempos de desarrollo, la rápida puesta en el mercado y los costos, maximizando al mismo tiempo la calidad de los productos de software resultantes. En particular, el reuso orientado a dominios se basa en el mismo concepto abocando a los mismos beneficios, pero considerando que dicha reutilización se generará dentro de dominios particulares. Dentro de esta área surge la Ingeniería de Líneas de Productos de Software (ILPS) que utiliza el reuso de dominios como mecanismo para lograr una estructura común de servicios (llamada similitudes), en la cual se basan todos los productos a ser desarrollados. A su vez, agrega una estructura variable (llamada variabilidad) que permite brindar mayor flexibilidad logrando crear productos con bases similares pero adaptados a las

necesidades específicas de los diferentes usuarios u organizaciones pertenecientes a esos dominios. En nuestro proyecto, nos basamos en la creación de nuevas técnicas y recursos para lograr una mejora dentro de un proceso integral de desarrollo de las LPSs. En particular nos enfocamos en la definición de actividades que diseñan y aplican artefactos de software en dos dominios específicos: el dominio geográfico y el de gobierno electrónico.

Palabras Clave: Líneas de Productos de Software - Artefactos de Software - GIS - Gobierno Electrónico

2. Contexto

La línea presentada se inserta en el contexto de los siguientes proyectos y acuerdos de cooperación:

- *Programa: Desarrollo de Software Basado en Reuso (04/F001)*. Acreditado por la Universidad Nacional del Comahue según Ordenanza 1268/13. Directora: Dra. Alejandra Cechich.
- *Proyecto: Reuso Orientado a Dominios*. Incluido dentro del Programa. Directora:

Dra. Agustina Buccella, Codirector: Mg. Juan Manuel Luzuriaga.

- *Acuerdo de Cooperación* entre el Laboratorio de Investigación en Ecología Bentónica y el Laboratorio de Parasitología e Histopatología de Moluscos del Instituto de Biología Marina y Pesquera Almirante Storni.

3. Introducción

La Ingeniería de Líneas de Productos de Software (ILPS) [4, 8, 18] contempla la definición de partes comunes de un conjunto de productos dentro de un dominio específico y al mismo tiempo provee un mecanismo para modelar la variabilidad dentro de los mismos. A diferencia del desarrollo de software individual, en la ILPS se debe realizar un mayor esfuerzo inicial para identificar y obtener aquellos requerimientos que formarán parte de la línea, es decir, la plataforma de los productos. Estos requerimientos son tanto funcionales como no funcionales, incluyendo particularmente los atributos de calidad de un sistema de software. Una vez obtenida la plataforma de la línea de productos, se deben identificar los requerimientos particulares de cada producto. Es en este punto, donde la flexibilidad de la plataforma juega un rol fundamental ya que es la que definirá, en cierta medida, el rango de productos que es posible crear desde la LPS. Es deseable que la base que contiene las características comunes sea lo más flexible posible para permitir el reuso de la misma, y lo suficientemente acotada para no generar un conjunto inmanejable de posibilidades.

Para generar o facilitar la flexibilidad de la plataforma es deseable que los módulos o componentes que la componen sean adaptables a los diferentes sistemas que puedan ser generados. Para lograr esto es necesario identificar y describir las funcionalidades de la línea de productos que pueden variar en función de las características que ofrecen, los requisitos que necesitan, o incluso en términos de la arquitectura subyacente. Es por esto que la plataforma de la LPS deberá proporcionar los medios para satisfacer las necesidades de diferentes clientes. Una forma de propiciar un ambiente para mejorar la flexibilidad de la LPS es por medio de la definición y uso de componentes bien definidos en base a los lineamientos de la Ingeniería de Software Basada en Componentes (ISBC) [3, 22]. También es necesaria la adopción o redefinición de una metodología para la gestión de variabilidad que

permita administrar la variabilidad durante todo su ciclo de vida [9, 21].

Dentro de las LPSs el primer aspecto a definir es el dominio en el cual los servicios y futuros componentes van a ser desarrollados para que puedan luego ser reusados en los productos creados para dicho dominio. En particular, dentro de la línea de investigación actual estamos trabajando en dos dominios específicos en donde ya contamos con algunos resultados visibles. Dichos dominios son:

- *Dominio Geográfico* [7, 11, 12]: Este dominio incluye el conjunto de datos, estructuras y operaciones necesarias para representar, manipular y visualizar la información geográfica. Construir una LPS dentro de este dominio significa explorar y definir las variabilidades posibles dentro de los subdominios que la información geográfica abarca. Esta división en subdominios más manejables se propuso en trabajos previos [5, 6] mejorando la complejidad inherente del dominio geográfico general. Por esto, se ha trabajado dentro del subdominio de ecología marina en el cual hemos desarrollado una taxonomía de servicios [6] basada en los estándares geográficos definidos por el Open Geospatial Consortium¹ (OGC) y la International Organization for Standardization² (ISO). En particular, dicha taxonomía se deriva de la especialización de la ISO/DIS 19119³ enfocándose en los servicios específicos del subdominio.

Al mismo tiempo hemos definido una metodología de desarrollo, la cual se basa en la construcción de artefactos de software para cada una de las actividades, comenzando con la taxonomía previamente descrita. En la Figura 1 podemos observar las actividades y artefactos de la metodología la cual se divide en dos tipos de análisis: de *dominio* y *organizacional*. El análisis de dominio involucra el análisis y diseño de la información dentro de un dominio específico pero enfocándose en una vista general. Luego, el análisis organizacional utiliza la organización de la información anterior para adaptarla al contexto de la LPS que está siendo desarrollada. Así, el análisis de dominio impacta directamente en el organizacional. En trabajos previos [1, 2, 13, 19, 20] hemos presentado las bases

¹<http://www.opengeospatial.org/>

²<http://www.iso.org>

³ISO/DIS 19119 : Geographic information Services, ISO/TS 2005

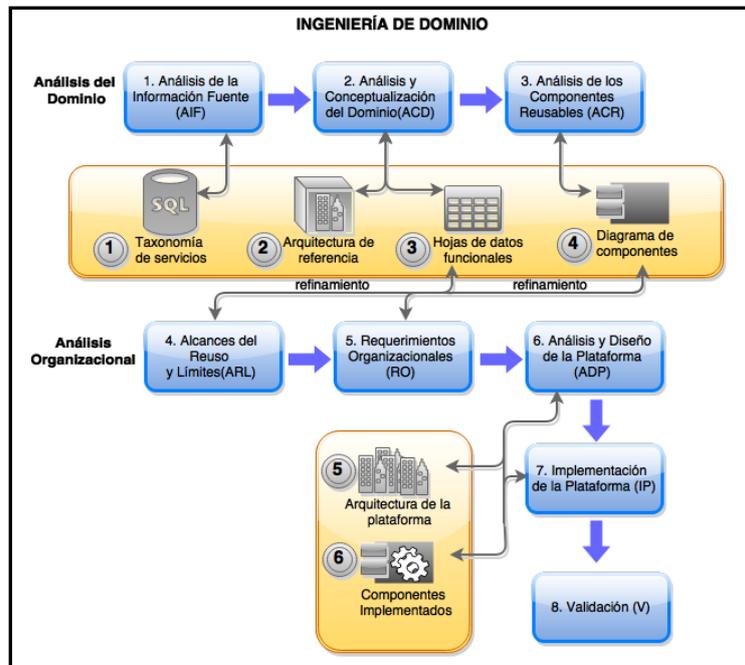


Figura 1. Actividades y artefactos de software involucrados en la actividad de ingeniería de dominio de una LPS

de la metodología dentro de la fase de ingeniería de dominio de las LPSs.

Así, en la Figura 1 podemos observar como primer artefacto a la taxonomía la cual es creada en la actividad de análisis de la información fuente (AIF). La misma clasifica los servicios posibles dentro de un dominio en particular. El segundo artefacto es la definición de una arquitectura de referencia como parte de la actividad de análisis y conceptualización del dominio (ACD). A su vez en esta actividad se debe crear el tercer artefacto definido: las hojas de datos funcionales. Éstas poseen cada una de las funcionalidades del dominio especificadas mediante una serie de ítems. Este artefacto es muy importante ya que posee todo el diseño de la interacción de los servicios para cumplir con una cierta funcionalidad. Es decir, incluye la definición de servicios comunes y variantes de acuerdo a las dependencias definidas para modelar e implementar la variabilidad [19, 20]. Por último, el cuarto artefacto es una estructura preliminar de componentes de software basados en la especificación previa de las hojas de datos funcionales. Dicha estructura debe crearse considerando la clase de servicios definidos (comunes, variantes y puntos variantes) y sus interacciones [1].

La misma se realiza en la actividad de análisis de los componentes reusables (ACR). Luego, en el segundo cuadro de la Figura 1 podemos observar las actividades y los artefactos de software (numerados del 5 al 7) creados para el análisis organizacional. Los artefactos 2 y 3 creados anteriormente son re-analizados y adaptados de manera de acercar la LPS a las aplicaciones específicas que serán derivadas. El desarrollo entonces de los artefactos 5 y 6, hojas de datos funcionales refinadas y componentes reusables respectivamente, serán efectuados como parte de las actividades alcances del reuso y límites (ARL) y requerimientos organizacionales (RO).

- *Domino de gobierno electrónico*⁴: Este dominio incluye el estudio de las tecnologías y las comunicaciones con el fin de ofrecer mejores servicios a los ciudadanos, optimizar la gestión pública, garantizar la transparencia, etc. En particular, la interoperabilidad es uno de los aspectos más estudiados en este dominio y para los cuales el reuso es una herramienta válida [10]. Es sabido que la interoperabilidad es uno de los problemas que hoy enfrentan los

⁴<http://www.jgm.gov.ar/sgp/paginas.dhtml?pagina=98>

estados, específicamente al querer compartir e intercambiar información para sus procesos de negocios. La misma dentro del estado es un nudo operacional si se quiere mejorar la eficiencia incidiendo directamente en tareas asociadas a una mejor atención al ciudadano. En algunos países se están implementando plataformas de interoperabilidad únicas dentro del estado, como es el caso de Chile, denominada Plataforma de Servicios Electrónicos del Estado (PISEE⁵). Así, se desprende la necesidad de una Ingeniería de Dominios cuidadosamente orientada a soportar reuso de servicios en pos de una interoperabilidad efectiva⁶. En este dominio se está desarrollando actualmente una taxonomía de servicios que permita definir y clasificar los servicios comunes y variables del mismo.

4. Líneas de investigación y desarrollo

Como se describió anteriormente, se han realizado varios avances con respecto al uso de un conjunto de servicios geográficos comunes y a la definición de artefactos y actividades que dan soporte al proceso de desarrollo de LPS. En este contexto hemos creado un proceso para el diseño e implementación de funcionalidades dentro de la LPS durante las actividades de la ingeniería de dominio. Aquí, los artefactos de software son creados y refinados de manera de diseñar funcionalidades que serán luego parte de la plataforma de la LPS. A su vez se han implementado herramientas de software que guían a los ingenieros de software en la definición, creación y aplicación de cada uno de los artefactos. Por ejemplo, en [2] hemos definido un proceso de selección asistida junto con su herramienta de soporte para reducir o eliminar las dificultades respecto a la complejidad inherente del uso de las taxonomías. Al mismo tiempo, hemos también implementado la herramienta *Datasheet Modeler* [13] la cual asiste a la tarea de creación de las hojas de datos funcionales en base a los servicios de la taxonomía permitiendo la aplicación de nuestro enfoque de variabilidad [20]. En la actualidad estamos refinando el enfoque para la gestión de la variabilidad tanto en el diseño como en la implementación y se están definiendo protocolos de derivación de componentes en tecnologías específicas de infor-

⁵http://wikigob.cl/mediawiki/index.php?title=Plataforma_Integrada_de_Servicios_Electr%C3%B3nicos_del_Estado

⁶Interoperability Solutions for European Public Administrations - <http://ec.europa.eu/isa/>

mación geográfica que permitan plasmar las restricciones del diseño previamente definidas.

Dentro del dominio del gobierno electrónico y en particular en los subdominios de las notificaciones, gestión de recursos y construcción de sitios accesibles se han realizado avances respecto a procesos normalizados basados en modelos de mejora y en estándares para incorporación de firma digital. También hemos diseñado un método para incluir accesibilidad web en etapas tempranas del diseño y una herramienta de soporte que utiliza estándares (pautas WCAG 1.0/2.01). Se ha profundizado en la evaluación automática de accesibilidad web a través de la clasificación de necesidades de usuarios no videntes mediante el uso de agentes inteligentes. Estos trabajos han sido publicados en [14, 15, 16, 17]. Dentro de este dominio se continuará con el desarrollo de técnicas para la mejora de la madurez de gobierno electrónico incluyendo sus prácticas más características, por ejemplo, la interoperabilidad basada en dominios.

5. Resultados y Objetivos

El objetivo principal de la línea de investigación es *Desarrollar técnicas y herramientas que mejoren el desarrollo de software basado en reuso a nivel de dominios, considerando los dominios específicos de los sistemas de información geográficos y las aplicaciones para gobierno electrónico*. Dentro de los nuevos avances, se han publicado artículos y terminado varios trabajos de fin de carrera los cuales se han enfocado en la instanciación y validación de la LPS creada para dos organizaciones que trabajan en el dominio. En particular se ha finalizado una tesis de grado denominada *Representación de variabilidad dentro de una Línea de Productos de Software aplicada al subdominio de Ecología Marina* del alumno Victor Pincheira con dirección de Agustina Buccella y se encuentran en ejecución dos tesis de grado sobre una temática similar. A su vez, en el dominio de accesibilidad se está finalizando una tesis de grado denominada *Generación automática de perfiles estereotipados para usuarios con discapacidad visual* del alumno Jerónimo Perfumo y dirigida por Rafaela Mazalú.

6. Formación de Recursos Humanos

El proyecto reúne aproximadamente a 10 investigadores, entre los que se cuentan docentes y alumnos de UNComa, y asesores externos. A su vez, el proyecto cuenta actualmente con 3

doctores, y un magister. Varios de los docentes-investigadores de GIISCo-UNComa han terminado o se encuentran próximos a terminar carreras de postgrado, uno de ellos ha finalizado recientemente su Beca CONICET Tipo II y otros dos poseen becas CONICET-Tipo I.

Referencias

- [1] M. Arias, A. Buccella, and A. Cechich. Towards semi-automatic component derivation from an spl variability model. In *CONAISI 2015: 3er Congreso Nacional de Ingeniería Informática/Sistemas de Información*, Buenos Aires, Argentina, 2015.
- [2] M. Arias, A. de Renzis, A. Buccella, A. Cechich, and A. Flores. Búsqueda de servicios para asistir en el desarrollo de una línea de productos de software. In *16 Simposio Argentino de Ingeniería de Software (ASSE'15) en el marco de las 44 Jornadas Argentinas de Informática (JAIIO)*, Rosario, Argentina, 2015.
- [3] L. Baum, M. Becker, L. Geyer, A. Gilbert, G. Molter, and V. Tamara. Supporting component-based software development using domain knowledge. In *Proceedings of the SCI 2000 Conference*, 2000.
- [4] J. Bosch. *Design and use of software architectures: adopting and evolving a product-line approach*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 2000.
- [5] A. Buccella, A. Cechich, M. Arias, M. Pol'la, S. Doldan, and E. Morsan. Towards systematic software reuse of gis: Insights from a case study. *Computers & Geosciences*, 54(0):9 – 20, 2013.
- [6] A. Buccella, A. Cechich, M. Pol'la, M. Arias, S. Doldan, and E. Morsan. Marine ecology service reuse through taxonomy-oriented SPL development. *Computers & Geosciences*, 73(0):108 – 121, 2014.
- [7] P. Burrough and R. McDonnell. *Principles of Geographical Information Systems*. Oxford University Press, 1998.
- [8] P. C. Clements and L. Northrop. *Software Product Lines : Practices and Patterns*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.
- [9] M. Galster, D. Weyns, D. Tofan, B. Michalik, and P. Avgeriou. Variability in software systems - a systematic literature review. *IEEE Transactions on Software Engineering*, 40(3):282–306, 2014.
- [10] G. Laskaridis, K. Markellos, P. Markellou, A. Panayiotaki, E. Sakkopoulos, and A. Tsakalidis. E-government and interoperability issues. *International Journal of Computer Science and Network Security*, 7(9):28–38, 2007.
- [11] P. Longley, M. Goodchild, D. Maguire, and D. Rhind. *Geographic Information Systems and Science*. John Wiley & Sons, 2001.
- [12] M.A. Rodríguez Luaces. *A Generic Architecture for Geographic Information Systems*. PhD thesis, Univerdade da Coruña, 2004.
- [13] M. Mancuso, A. Buccella, A. Cechich, M. Arias, and M. Pol'la. Datasheet modeler: Una herramienta de soporte para el desarrollo de funcionalidades en líneas de productos de software. In *XXI Congreso Argentino en Ciencias de la Computación*, Junin, Argentina, 2015.
- [14] R. Mazalú and A. Cechich. Soporte inteligente para la evaluación de accesibilidad web relacionada con limitaciones visuales. In *16 Simposio Argentino de Ingeniería de Software (ASSE'15) en el marco de las 44 Jornadas Argentinas de Informática (JAIIO)*, pages 205–219, Rosario, Argentina, 2015.
- [15] R. Mazalu, A. Cechich, and A. Martin. Automatic profile generation for visual - impaired users. In *Proceedings of the Argentine Symposium on Software Engineering (ASSE)*, Córdoba, Argentina, 2013.
- [16] R. Mazalu, A. Cechich, and A. Martin. Evaluación de accesibilidad del contenido web utilizando agentes. In *X Workshop Ingeniería de Software (WIS)*, Mar del Plata, Argentina, 2013.
- [17] G. Miranda, A. Martin, R. Mazalu, G. Gaetan, and V. Salda no. Evaluación de accesibilidad del contenido web utilizando agentes. In *X Workshop Ingeniería de Software (WIS)*, Mar del Plata, Argentina, 2013.
- [18] Klaus Pohl, Günter Böckle, and Frank J. van der Linden. *Software Product Line Engineering: Foundations, Principles and Techniques*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [19] M. Pol'la, M. Arias, A. Buccella, and A. Cechich. Un sistema de anotaciones para la especificación de componentes de una línea de productos de software. *Revista Tecnológica y Ciencia de la Universidad Tecnológica Nacional*, 6(1):116–122, 2015.
- [20] M. Pol'la, A. Buccella, A. Arias, and A. Cechich. Sevatax: Service taxonomy selection & validation process for spl development. In *XXXIV International Conference of the Chilean Society of Computer Science (SCCC'15)*, Santiago, Chile, 2000. IEEE Computer Society Press.
- [21] M. Sinnema and S. Deelstra. Classifying variability modeling techniques. *Information and Software Technology*, 49:717–739, July 2007.
- [22] Clemens A. Szyperski. *Component software - beyond object-oriented programming*. Addison-Wesley-Longman, 1998.

Software para la gestión de requerimientos del Modelo Conceptual de un sistema de información

Oscar Carlos Medina, Marcelo Martín Marciszack, Mario Alberto Groppo,
Castro Claudia, Moreno Juan Carlos, Moyano Enrique Humberto

GIDTSI, Grupo de Investigación, Desarrollo y Transferencia de Sistemas de
Información

Universidad Tecnológica Nacional, Facultad Regional Córdoba
Maestro López esq. Av. Cruz Roja Argentina, Ciudad Universitaria – (5016) Córdoba
oscarmedina@gmail.com, marciszack@gmail.com, sistemas@groppo.com.ar,
ingclaudiacastro@gmail.com, jmoreno33@gmail.com, enriquemoyano@gmail.com

Resumen

El presente trabajo describe una aplicación web denominada SIAR (Sistema Integral de Administración de Requerimientos) que administra y valida los requerimientos del Modelado Conceptual de un sistema de información mediante Casos de Uso, según los lineamientos de UML (Lenguaje Unificado de Modelado). Los Casos de Uso son útiles para la generación y análisis de requisitos de sistemas. La finalidad principal de SIAR es la administración de Casos de Uso con una herramienta informática que agilice su registración, normalice su contenido y posibilite implementar validaciones funcionales, como por ejemplo un método automatizado de análisis de consistencia de Casos de Uso, para lo cual el sistema genera un grafo con la transición de estados de cada Caso de Uso, expresado en el protocolo XPDL (Lenguaje de Definición de Flujo de Trabajo), que es analizado en un simulador de autómatas finitos deterministas para verificar la cohesión de los escenarios en él definidos.

Palabras clave: Caso de Uso, Requerimientos, UML, XPDL, Autómata finito determinista.

Contexto

SIAR (Sistema Integral de Administración de Requerimientos) es la denominación que se le dio al software que gestiona los requerimientos funcionales de un sistema de información según los lineamientos de UML (Lenguaje Unificado de Modelado).

Esta aplicación se originó dentro del proyecto de investigación “Validación de Requerimientos a través de Modelos Conceptuales” del GIDTSI (Grupo de Investigación, Desarrollo y Transferencia de Sistemas de Información), dependiente del Departamento Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, Facultad Regional Córdoba que “busca dar solución a uno de los principales problemas de la Ingeniería de Requisitos relacionado a la elicitación y especificación de requerimientos, que vincula las distintas etapas del proceso de desarrollo de software manteniendo la trazabilidad de los mismos hasta su validación e implementación”.

Introducción

Como las actividades principales del Modelo Conceptual son la registración y estandarización de requerimientos funcionales, la construcción de SIAR tiene por fin administrar en forma integral los requerimientos funcionales de un proyecto de software según la metodología UML. SIAR es una aplicación web que permitió registrar en forma normalizada los casos de uso y cuya primera versión comprende el siguiente alcance:

- Administración de los atributos de un proyecto de software y su versionado.
- Diseño y validación del Modelo Conceptual.
- Gestión de los alcances de cada versión del proyecto y los casos de uso asignados.
- Administración de los atributos de un caso de uso, incluyendo actores, pre-condiciones, post-condiciones, cursos de acción, normal y alternativos, y su versionado.
- Clasificación, priorización y trazabilidad de los casos de uso.
- Visualización de consultas y generación de reportes en distintos formatos, inclusive XPDL, para comunicarse con otras aplicaciones.
- Gestión de atributos de procesos de negocio, de actividades de negocio que los componen y los casos de uso asociados a estas actividades.
- Registración y consulta de un glosario por proyecto, con entradas y sinónimos, siguiendo las recomendaciones de LEL (Léxico Extendido del Lenguaje), que es una estrategia de modelado de requisitos basada en lenguaje natural.

Líneas de Investigación, Desarrollo e Innovación

a) Construcción de un software para la gestión integral de Casos de Uso.

SIAR permite trabajar con diferentes Sistemas, cada uno con sus propias Versiones, cada Versión cubre un número limitado de Alcances, cada Alcance gestiona un grupo de Casos de Uso, cada Caso de Uso está compuesto por una secuencia ordenada de Pasos, finalmente un paso puede o no tener Alternativas.

Los Pasos de un Caso de Uso permiten efectuar tareas que son descriptos con el soporte del sistema que se ha desarrollado.

Desde el punto de vista conceptual un Caso de Uso es la descripción de una secuencia de interacciones entre el sistema y uno o más Actores. Las personas o entidades que participarán en un caso de uso se denominan Actores.

Cada caso de uso tiene pre-condiciones que se deben cumplir para que el flujo de eventos se pueda llevar a cabo. Dichas pre-condiciones son las reglas o condiciones que se deben cumplir antes de que sea iniciado el caso de uso.

También posee post-condiciones que reflejan el estado en que se queda el Sistema una vez ejecutado el caso de uso.

Puede clasificarse con diferentes niveles de Complejidad según el criterio de cada Analista funcional en Muy Alta, Media y Baja.

A su vez, un Caso de Uso puede ser del tipo Concreto o Abstracto. Un caso de uso es Abstracto si no puede ser realizado por sí mismo, o si no es Concreto ya que puede ser iniciado por un actor y realizado por sí mismo.

Los pasos son las actividades que deberán realizarse para llevar a cabo algún proceso (Rumbaugh, Jacobson, Booch, 1999).

Cada paso posee alternativas que a su vez puede tener pasos y estos volver a tener alternativas.

b) Alcances de la versión inicial.

El aplicativo busca cubrir las siguientes necesidades:

b.1) Gestión de requerimientos: Se planteó la necesidad de generar un sistema que no solo administre los casos de usos por separado, sino también gestione en forma integral todo el proyecto que lo contiene. Es decir, sistema con sus versiones, por cada versión del sistema sus alcances y por cada alcance sus casos de usos con sus versiones. Para cumplir con esta funcionalidad se incluyó también el concepto de sistema, versiones y alcances de versiones. A su vez cada caso de uso incluye pasos y alternativas. También se propone hacer una trazabilidad de los cambios.

b.2) Configuración del entorno de desarrollo.

b.3) Interfaz de usuario.

b.4) Administración de proyectos de sistemas.

b.5) Administración de Casos de Uso.

b.6) Versionado de Casos de Uso.

b.7) Exportación a archivos XML: Esta es una funcionalidad que proporciona el sistema para poder intercambiar datos con otras aplicaciones por ejemplo un autómata finito. El archivo XML representa en este protocolo de intercambio de datos el grafo de estados del Caso de Uso que surge de la equivalencia con sus pasos y alternativas.

b.8) Consultas.

Es posible obtener las siguientes salidas del sistema:

- Sistemas
- Versionado de Sistemas
- Alcances por Versión de Sistema
- Casos de Uso por Alcance
- Versionado de Casos de Uso

- Comparación entre dos Versiones de un Caso de Uso

- Reportes de Casos de Uso

Para la versión vigente de un Caso de Uso se pueden imprimir los siguientes reportes:

- Conversión a archivo XML para ingresar a Autómata Finito Determinista.

- Gráfico de estructura del Caso de Uso en forma de árbol.

- Reporte PDF del Caso de Uso.

- Tabla de Estados del Caso de Uso.

- Información de Usuarios.

c) Validación de consistencia de Casos de Uso con simuladores de autómatas finitos.

En la actualidad UML es reconocido como el estándar para el modelado de proyectos de software orientado a objetos.

Los componentes principales de esta metodología son los Casos de Uso ya que especifican el comportamiento deseado del sistema. Por definición el caso de uso “especifica una secuencia de acciones, incluyendo variantes, que el sistema puede ejecutar y que produce un resultado observable de valor para un particular actor”.

Una vez generado un caso de uso, es necesario comprobar y asegurar su validez. La verificación de consistencia de la secuencia de acciones descrita en el caso de uso es una de las tareas que permiten su validación.

Es deseable que esta verificación pueda realizarse de manera automatizada para lo cual se podría trabajar con autómatas finitos deterministas, ya que un autómata finito es un conjunto de estados y un control que se mueve de un estado a otro en respuesta a entradas externas (Marciszack, Pérez, Castro, 2013). Se llama determinista al autómata que puede estar únicamente en un estado en un

momento determinado. El desafío es transformar el caso de uso en un autómata finito determinista para validar su cohesión secuencial.

Partiendo de estas premisas, el aporte del presente análisis a la validación de requerimientos que se propone lograr el proyecto de investigación, consiste en dar respuesta a la siguiente pregunta:

¿Es factible determinar un método, basado en simuladores de autómatas finitos deterministas, que verifique la consistencia de la secuencia de acciones, incluyendo variantes, definidas en un caso de uso?

La funcionalidad de SIAR que se detalla a continuación propone una alternativa viable para esta necesidad, que se realiza en 3 pasos:

c.1) Registración normalizada del requerimiento.

c.2) Transformación del Caso de Uso en una máquina de estados.

c.3) Validación de la consistencia secuencial de los cursos de acción del Caso de Uso.

Resultados y Objetivos

Se pone a consideración el aplicativo SIAR que registra, normaliza y transforma un caso de uso a formato XPDL y un simulador de autómatas finitos, que permite verificar la consistencia secuencial de los distintos caminos del caso de uso.

El aporte que realiza SIAR al proyecto en el que fue concebido, “Validación de Requerimientos a través de Modelos Conceptuales”, es el de constituirse como plataforma de software integradora de las aplicaciones que se utilizan en cada una de las líneas de investigación.

El presente desarrollo está fundamentado en una “Propuesta Metodológica para validación de

Requerimientos Funcionales a través de Modelos Conceptuales” registrada en la República Argentina con Derecho de autor de producciones tecnológicas (rubro “Modelo de organización y/o gestión, Ciencia y cultura-Ciencia y tecnología”) según Expediente No.5229942. Esta metodología expone cómo se llega a identificar la necesidad de construir SIAR y la especificación de requerimientos funcionales de sus alcances, módulos y funcionalidades.

También “S.I.A.R. – Sistema Integral de Administración de Requerimientos” está registrado en la República Argentina con Derecho de autor de producciones tecnológicas (rubro “Máquina, equipo, instrumento y/o herramienta o su/s componente/s. Informática-software. Ciencia y cultura-Ciencia y tecnología”) según Expediente No.5229955.

En lo que respecta a la funcionalidad de análisis de consistencia, SIAR ofrece un método automatizado para validar la cohesión de un caso de uso desde el punto de vista de la transición de estados definidos intrínsecamente en los pasos de su especificación funcional. Permitiendo también enlazar este proyecto con un trabajo académico de la carrera Ingeniería en Sistemas de Información, el del Grupo de Herramientas Didácticas de Informática Teórica que contribuyó con el simulador de autómatas finitos.

Finalmente se tiene previsto en el año 2016 hacer una transferencia del software a una empresa especializada en desarrollos de seguridad informática de la región en el marco de un programa de extensión universitaria para la generación y transferencia de conocimientos.

Formación de Recursos Humanos

En el marco del actual proyecto de investigación, en el 2015 se doctoró en la

Universidad de Vigo, Marcelo Marciszack bajo la dirección del Dr. Manuel Pérez Cota con tesis que abordó el tema de "Validación de especificaciones funcionales de esquemas conceptuales a través de abstracciones"; y está previsto la realización de las siguientes tesis y trabajos finales de carreras: título de Especialista en Ingeniería en Sistemas de Información en la U.T.N. F.R.C. de la Ing. Claudia Castro cuyo tema de trabajo es "Un modelo conceptual para la obtención del sistema de Información a partir del Sistema del Negocio"; Tesis final dentro de la Maestría en Ingeniería en Sistemas de Información del tesista Juan Carlos Moreno cuyo tema es "Desarrollo de un marco teórico metodológico de técnicas de testing de usabilidad para verificar especificaciones no funcionales" cuyo director de Tesis es el director del presente proyecto de investigación; y la postulación de Tesis de Doctorado en la U.T.N. F.R.C. de Oscar Carlos Medina bajo la dirección del Dr. Mario Groppo cuyo tema es "Metodología para implementación de patrones en buenas prácticas de e-Gobierno". Además, se incorporan tres becarios alumnos de investigación, un becario Graduado BINID, que colaboran en la investigación bibliográfica, diseño y construcción de la herramienta informática. Al mismo tiempo, y como contribución a la formación de los integrantes del proyecto, se elaborarán informes técnicos, como así también artículos para ser publicados en Congresos, Conferencias y reuniones científicas tanto a nivel nacional como a nivel internacional.

Referencias

Brooks, Frederik P. (1987). No Silver Bullet. Essence and Accidents in Software Engineering. IEEE Computer.

Chakraborty, Samarjit (2003). Formal Languages and Automata Theory-Regular Expressions and Finite Automata-. Computer Engineering and Networks Laboratory Swiss Federal Institute of Technology (ETH) Zurich.

Davis, A. (1993). Software requirements. Object, functions and states. Prentice Hall international Inc.

Jacobson, Ivar y otros (1992). Object Oriented Software Engineering. A Use Case Driven Approach. Addison Wesley.

Leonardi, C., Leite, J.C.S., Rossi, Gustavo (2001). Una estrategia de Modelado Conceptual de Objetos, basada en Modelos de requisitos en lenguaje natural. Tesis de Maestría Universidad Nacional de la Plata.

Pérez, J. D. (2007). Notaciones y lenguajes de procesos. Una visión global. Tesis de Doctorado Universidad de Sevilla.

Rumbaugh, J., Jacobson, I., Booch, G. (1999). The Unified Modelling Language Reference. Addison Wesley.

Sommerville, I. (2005). Software Engineering, Computing Department, Lancaster University, John Wiley & Sons Ltd.

Antecedentes de esta investigación:

Marciszack, Marcelo, Pérez, Ramiro, Castro, Claudia (2013). Validación de Requerimientos a través de Modelos Conceptuales – Modelos y Transformaciones. WICC 2013.

Marciszack, Marcelo, Pérez Cota, Manuel, Groppo, Mario Alberto (2015). Metodología y herramienta de soporte para validar modelos conceptuales a través de máquinas abstractas. Universidad de Palermo. Revista de Ciencia y Técnica.

U.T.N. F.R.C. (2009). Proyecto Construcción de Herramientas Didácticas para la enseñanza y ejercitación práctica en laboratorio de Informática Teórica en las Carreras con Informática. Manual de Usuario – Grupo de Herramientas Didácticas.

Modelo Integral para la Evaluación de la Calidad de la Accesibilidad al Contenido Web

Cecilia Gallardo¹, Ana Funes², Hernán Ahumada¹

¹ Facultad de Tecnología y Ciencias Aplicadas
Universidad Nacional de Catamarca
Maximio Victoria 55 – 4700 Catamarca, Argentina
{ceciliagallardo, hcahumada}@tecno.unca.edu.ar

² Departamento de Informática
Facultad de Ciencias Físico-Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950 - 5700 San Luis, Argentina
afunes@unsl.edu.ar

Resumen

El presente trabajo, que se enmarca en el área del Aseguramiento de la Calidad del Software, busca dar respuesta a la necesidad de contar con un modelo integral o marco de referencia que permita conceptualizar y, posteriormente, evaluar la accesibilidad al contenido web.

Para tal fin, se propone, por un lado, el desarrollo de un modelo de conceptos de calidad correspondiente a la subcaracterística *Accesibilidad* del modelo de calidad de la norma ISO 25010, para lo cual se analizarán las Pautas de Accesibilidad al Contenido Web (WCAG) 2.0. El proceso completo de medición y evaluación se implementará siguiendo los lineamientos de la estrategia integrada de Medición y Evaluación GOCAME (Goal-Oriented Context-Aware Measurement and Evaluation). Otro objetivo de este trabajo es el desarrollo, a partir del modelo de conceptos de calidad, de un modelo cuantitativo, basado en el método LSP

(Logical Scoring of Preference), que permita evaluar la calidad global de las aplicaciones web en cuanto a la accesibilidad.

Palabras clave: accesibilidad web, métricas, calidad web.

Contexto

El presente trabajo de investigación se encuentra enmarcado en una colaboración entre investigadores del Proyecto de Incentivos código 22/F222 *Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software*, de la Universidad Nacional de San Luis, y docentes de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca.

En particular, esta propuesta de investigación se desarrolla como una tesis de posgrado de la carrera *Maestría en*

Introducción

La Ingeniería Web, basada en la Ingeniería de Software, comprende el uso de principios científicos, de ingeniería y de enfoques disciplinados y sistemáticos para la especificación, implementación, operación y mantenimiento de Aplicaciones Web de alta calidad, para satisfacer con éxito la diversidad de usuarios y los requerimientos únicos de este tipo de sistemas [1]. La Ingeniería Web consiste en un proceso, por lo que abarca una serie de áreas como ser: Calidad Web, Usabilidad Web, Métricas Web, Evaluación del desempeño de Sistemas Web, entre otras [2].

La definición e implementación de estándares fijan los atributos deseables del software de calidad, a la vez que surgen modelos y metodologías para la evaluación de la calidad. Mediante estándares como ISO/IEC 9126 [3] y su sucesor ISO/IEC 25010 [4] se ha tratado de determinar y categorizar los factores que afectan a la calidad del software. Por ejemplo, el estándar ISO/IEC 25010 define un conjunto de características, así como las relaciones entre ellas, que sirven de base para la especificación de requisitos de calidad y evaluación de la misma. Uno de los modelos propuestos es el Modelo de Calidad de Sistema/Producto de Software, que abarca cualidades internas y externas, que se relacionan con propiedades estáticas y dinámicas de un sistema software. En esta nueva versión del estándar, se ha reformulado la definición de la característica Usabilidad, agregando dos sub-características: Protección contra errores de usuario y Accesibilidad. A su

vez, la Accesibilidad es definida, en este estándar, como “grado en que un producto o sistema puede ser utilizado por personas con la más amplia gama de características y capacidades para lograr un objetivo determinado en un contexto de uso especificado”.

Existen también, propuestas de modelos de calidad específicos que se ajustan a las particularidades de los requerimientos de calidad web, tal como el framework de modelado de calidad web denominado 2Q2U (Quality, Quality in Use, actual Usability and User experience), presentado como una extensión de los modelos de calidad de ISO 25010 [5] [6].

Desde otro contexto y con un enfoque práctico, el Consorcio de la World Wide Web [7], considera a la Accesibilidad Web como un requerimiento importante a ser tenido en cuenta en el desarrollo de aplicaciones web, con el objetivo de que todas las personas, principalmente aquellas con discapacidad, puedan percibir, entender, navegar, interactuar con la Web, y también poder contribuir a la misma.

Entre el conjunto de los componentes del desarrollo Web, que deben interactuar para que la Web sea accesible, se encuentra el “contenido web”, el cual se refiere a la información dentro de una página web, como ser: texto, imágenes y sonidos, código o marcado, etc. Para ayudar a hacer el contenido web más accesible, la iniciativa WAI (Web Accessibility Initiative) del W3C desarrolló las “Pautas de Accesibilidad para el Contenido Web (WCAG)” [8], las cuales hoy constituyen un estándar internacional y proporcionan pautas, criterios de éxito comprobables y técnicas, los cuales se utilizan para evaluar los requerimientos de

accesibilidad web de acuerdo a las necesidades de diferentes grupos y situaciones.

Teniendo en cuenta lo anteriormente expuesto, se concluye que la Accesibilidad no solo representa una subcaracterística importante dentro del Modelo de Calidad de Sistema/Producto de Software del estándar ISO sino que también es considerada un componente esencial dentro del consorcio W3C. Sin embargo, resulta necesario extender y definir un modelo de calidad específico en el que se incluyan atributos medibles para conceptualizar la accesibilidad web, debido a que tanto en el modelo de ISO/IEC 25010 como en el modelo propuesto por 2Q2U, sólo se considera a la Accesibilidad como una subcaracterística de Usabilidad, no especificando propiedades de bajo nivel medibles.

Por otra parte, para realizar la evaluación de la calidad de un producto software, no basta solamente con la definición de modelos de calidad. Estos modelos deben ser instanciados, es decir, se deben considerar las características, atributos y relaciones que sean relevantes para un propósito dado y una necesidad de información de una categoría de entidad en concreto [9]. Entonces, para impulsar un proceso de medición y evaluación de la calidad acorde a los requerimientos, se debe adoptar una estrategia específica. En este sentido, consideramos que la estrategia integrada de Medición y Evaluación GOCAME (Goal-Oriented Context-Aware Measurement and Evaluation) [9] representa una alternativa adecuada para tal fin, ya que se trata de una estrategia multipropósito que sigue un enfoque orientado a objetivos, sensible al contexto y centrada en la necesidad de información de una organización. Soporta de manera

simultánea tres capacidades [10]: un marco conceptual; especificaciones de proceso de medición y evaluación desde diferentes puntos de vista; y la metodología WebQEM (Web Quality Evaluation) [11].

Entre las actividades que plantea la metodología WebQEM, se encuentra el diseño de indicadores parciales/globales para obtener el grado de cumplimiento global de las características de alto nivel que se estén evaluando. Para realizar esto, WebQEM utiliza el método LSP (Logical Scoring of Preference), el cual es un método cuantitativo basado en técnicas de puntuación y lógica continua de preferencias propuesto por Dujmovic [12] [13] [14]. Básicamente, LSP permite establecer criterios de evaluación, especificando las propiedades esperadas de un sistema. En este punto, todos los valores de los indicadores elementales pueden agruparse adecuadamente diseñando una estructura de agregación por niveles, que permite obtener una preferencia (valor de indicador global/parcial) de acuerdo a las necesidades del usuario y el punto de vista de la evaluación.

Líneas de Investigación, Desarrollo e Innovación

Los ejes principales de la presente línea de investigación son:

- Conceptualización de la accesibilidad al contenido web, analizando y definiendo las propiedades específicas que deben ser tenidas en cuenta en un contexto en particular.
- Medición de la accesibilidad al contenido web, estableciendo los métodos de cálculo o

procedimientos en forma de métricas, para obtener un valor numérico por cada atributo a ser considerado.

- Evaluación de la accesibilidad al contenido web, definiendo diferentes criterios de evaluación, para obtener el valor de un indicador global del Sistema/Producto evaluado.

Resultados y Objetivos

El principal objetivo de este trabajo es elaborar un modelo de requerimientos de calidad y un modelo cuantitativo, que permitan conceptualizar, medir y evaluar el grado de cumplimiento de la Accesibilidad al Contenido Web de manera integral.

Objetivos Específicos

- Definir los requerimientos de calidad para la Accesibilidad al Contenido Web, sobre la base del análisis e identificación de técnicas y herramientas proporcionadas por las Pautas de Accesibilidad al Contenido Web (WCAG) 2.0. y tomando en cuenta las necesidades de los usuarios del contexto de aplicación web elegido para la evaluación.
- Construir un catálogo de métricas que permitan medir los atributos hojas del árbol de requerimientos definido, analizando y considerando las métricas existentes que se adecúen a los requisitos definidos, o bien, diseñando métricas propias.
- Desarrollar un modelo cuantitativo de evaluación, aplicando el método

LSP, que permita obtener los valores de indicadores parciales de las características a diferentes grados de abstracción del árbol de requerimientos planteado en este trabajo, para conseguir finalmente el valor de un indicador global del Sistema/Producto evaluado.

Formación de Recursos Humanos

El presente trabajo es la base para el desarrollo de una tesis de posgrado de la carrera de posgrado “Maestría en Ingeniería de Software” de la Universidad Nacional de San Luis. La misma está siendo llevada adelante por la Licenciada en Sistemas de Información, Cecilia Gallardo, docente de la Universidad Nacional de Catamarca. Dicha tesis es dirigida por la Magister Ana Funes, docente e investigadora de la Universidad Nacional de San Luis, y a su vez, codirigida por el Dr. Hernán Ahumada, docente e investigador de la Universidad Nacional de Catamarca.

El trabajo de tesis se encuentra en la etapa inicial de escritura del informe, concretamente, en la confección del Marco Teórico.

Referencias

- [1] S. Murugesan, Y. Deshpande, S. Hansen y A. Ginige, «Web Engineering: A New Discipline for Development of Web-based Systems,» de *Proceedings of the First International Conference of Software Engineering (ICSE) Workshop on Web Engineering*, Los Angeles, USA, 1999.

- [2] S. Murugesan, «Web Application development: Challenges and the role of Web Engineering,» de *Web Engineering: Modelling and Implementing Web Applications (Human-Computer Interaction Series)*, London, Springer, 2008, pp. 7-32.
- [3] *ISO/IEC 9126*, 1991.
- [4] *ISO/IEC 25010*, 2011.
- [5] P. Lew, L. Olsina y L. Zhang, «Quality, Quality in Use, Actual Usability and User Experience as Key Drivers for Web Application Evaluation,» de *10th Int'l Congress on Web Engineering (ICWE2010)*, Vienne, Austria, 2010.
- [6] L. Olsina, P. Lew, A. Dieser y B. Rivera, «Updating Quality Models for Evaluating New Generation Web Applications,» de *Journal of Web Engineering*, 2012.
- [7] W3C, «World Wide Web Consortium,» [En línea]. Available: <https://www.w3.org/>. [Último acceso: 01 Febrero 2016].
- [8] W3C, «Web Content Accessibility Guidelines (WCAG) 2.0,» Diciembre 2008. [En línea]. Available: <http://www.w3.org/TR/WCAG20/>. [Último acceso: 01 Febrero 2016].
- [9] L. Olsina, F. Papa y H. Molina, «How to Measure and Evaluate Web Applications in a Consistent Way,» de *Web Engineering: Modelling and Implementing Web Applications*, London, Springer-Verlag, 2008, pp. 385-420.
- [10] P. Becker, F. Papa y L. Olsina , «Enhancing the Conceptual Framework Capability for a Measurement and Evaluation Strategy,» de *Current Trends in Web Engineering*, Springer, 2013, pp. 104-116.
- [11] L. Olsina y G. Rossi , «Measuring Web Application Quality with WebQEM,» *IEEE Multimedia*, pp. 20-29 Vol. 9(4), 2002.
- [12] J. Dujmovic, «Continuous Preference Logic for System Evaluation,» de *In Proceedings of Eurofuse 2005*, Belgrade, 2005.
- [13] J. Dujmovic, «A Method for Evaluation and Selection of Complex Hardware and Software Systems,» de *The 22nd International Conference for the Resource Management and Performance Evaluation of Enterprise Computing Systems. CMG 96 Proceedings*, 1996.
- [14] J. Dujmovic, «Partial absorption function,» *Journal of the University of Belgrade, EE Dept., Series Mathematics and Physics*, pp. 156-163, 1979.

Accesibilidad de la Información en Sitios Web argentinos

Rossi Bibiana, Ortiz Claudia, Chapetto Viviana

Departamento de Ciencias Básicas, Universidad Nacional de Luján (UNLu)

brossi@unlu.edu.ar, cortiz@unlu.edu.ar, vchapetto@yahoo.com.ar

Resumen

En 2010, se sancionó la ley 26.653 de Accesibilidad de la Información, que requiere que los sitios web del sector público (entes públicos estatales y no estatales, empresas privadas concesionarios de servicios públicos y empresas prestadoras o contratistas de bienes y servicios) respeten las normas y requisitos de accesibilidad recomendados por la ONTI (Oficina Nacional de Tecnologías de la Información). El proyecto que se lleva a cabo en la Universidad Nacional de Luján se propone realizar una evaluación del estado actual de los sitios web argentinos para determinar el grado de adhesión a las Normas de Accesibilidad Web 2.0, reglamentadas en Disposición 2/2014.

Palabras clave: accesibilidad, accesibilidad web.

Contexto

Este trabajo de investigación se encuentra radicado en la Universidad Nacional de Luján, en su Departamento de Ciencias Básicas. El inicio fue en diciembre del 2014 y tiene como duración 4 años (2014-2018).

Introducción

La Accesibilidad Web es el acceso universal a la Web, “independientemente del tipo de hardware, software, infraestructura de red, idioma, cultura, localización geográfica y capacidades de los usuarios” (W3C, 2005). Internet se ha

transformado en el medio masivo de transmisión de información y acceso al conocimiento, esto ha generado la necesidad de medir y valorar qué tan asequible es el acceso para todas las personas. El diseño de los sitios web puede crear barreras para determinados grupos de usuarios, limitando así la igualdad de acceso a la información.

Para evaluar la accesibilidad de los sitios web se definieron parámetros a tener en cuenta en el diseño de los portales, y se propusieron aplicaciones que permiten realizar la evaluación en forma automática. El World Wide Web Consortium (W3C) lanzó en 1997 la Web Accessibility Initiative (WAI) para desarrollar y promover el uso de estrategias, pautas y recursos que buscan hacer accesible a la World Wide Web para personas con discapacidad. (W3C, 1997). La WAI publicó, en 1999, su primera guía de pautas de accesibilidad “Web Content Accessibility Guidelines 1.0 (WCAG)”, cuya segunda versión del 2008, alcanzó el nivel de estándar ISO (ISO/IEC 40500:2012). (WAI, 2013).

En Argentina en 2010, se sancionó la ley 26.653 de Accesibilidad de la Información en las Páginas Web. Esta ley requiere que los sitios web del sector público (entes públicos estatales y no estatales, empresas privadas concesionarios de servicios públicos y empresas prestadoras o contratistas de bienes y servicios) respeten las normas y requisitos recomendados por la ONTI. (Argentina, 2010). La ley 26.653 “reconoce la necesidad de facilitar el acceso

a los contenidos de las páginas Web a todas las personas con discapacidad con el objeto de garantizarles la igualdad real de oportunidades y trato, evitando así todo tipo de discriminación” para “garantizar el derecho a la libertad de expresión y opinión, incluida, la libertad de recabar, recibir y facilitar información e ideas, en igualdad de condiciones mediante cualquier forma de comunicación” y que “pueda ser comprendida y consultada por usuarios que posean diversas configuraciones en su equipamiento o en sus programas”.

En 2011, se aprueba la Guía de Accesibilidad 1.0 para Sitios Web del Sector Público Nacional como parte de los Estándares Tecnológicos para la Administración Pública Nacional (ETAPS). La Guía “contiene pautas de accesibilidad de la información, que facilitan el acceso a los contenidos de la Web a todas las personas con discapacidad y a todos los usuarios, cualquiera sea la aplicación que utilicen (navegador de sobremesa, navegador de voz, teléfono móvil, PC de automóvil, etc.) o las limitaciones del entorno (ambientes ruidosos, habitaciones infra o supra iluminadas, manos libres, etc.), con la finalidad de garantizarles la real igualdad de trato y oportunidades, evitando así todo tipo de discriminación”. (ONTI, 2011)

En agosto de 2014, ONTI publica la disposición N° 2, en la que se aprueba la Norma de Accesibilidad Web 2.0, en vigencia actualmente, que establece los requisitos para contenidos web basándose en las Pautas de Accesibilidad para el Contenido Web (WCAG) en su versión 2.0 publicadas en el año 2008 por el W3C (ONTI, 2014). El objetivo principal es establecer criterios que permitan tanto desarrollar contenidos accesibles como evaluar la accesibilidad de contenidos existentes. Al establecer las características

que deben cumplir los contenidos disponibles mediante tecnologías web, se ha tenido en cuenta la mayoría de los tipos de discapacidad (visuales, auditivas, físicas, del habla, cognitivas, del lenguaje, de aprendizaje y neurológicas) y también las necesidades de las personas de edad avanzada.

Las WCAG 2.0, en las cuales se basa la Norma 2.0 vigente en Argentina, están organizadas en 4 niveles: principios, pautas, criterios de conformidad y técnicas.

- **Principios generales**, indican que todo contenido deber ser perceptible, operable, comprensible y robusto.
- **Pautas**, los 4 principios se distribuyen en 12 pautas que sirven de marco para comprender los criterios y las técnicas.
- **Criterios de conformidad**, cada pauta posee criterios, redactados en forma de enunciados verificables sobre el contenido web. Cada uno de los criterios está asociado a un nivel de conformidad:
 - ✓ **Nivel A**: es el nivel mínimo, la página web satisface todos los criterios de conformidad de nivel A, o se proporciona una alternativa conforme al nivel A.
 - ✓ **Nivel AA**: la página web satisface todos los criterios de conformidad de nivel A y AA, o se proporciona una alternativa conforme al nivel AA.
 - ✓ **Nivel AAA**: la página web satisface todos los criterios de conformidad de nivel A, AA y AAA, o proporciona una alternativa conforme al nivel AAA.
- **Técnicas**, para cada pauta y criterio, hay técnicas documentadas, aplicables al desarrollo o evaluación de contenidos web. Estas técnicas son informativas, no obligatorias, y se dividen en tres categorías: técnicas suficientes, técnicas recomendables y fallos comunes.

En la disposición 2/2014 de la ONTI, se aprueban los Niveles Mínimos de Conformidad, que comprenden 12 pautas y 35 criterios (todos de Nivel A).

En una lectura preliminar, se identificaron distintas publicaciones sobre el acceso a los sitios web.

- En 2008, se analizaron 188 bibliotecas universitarias, 66 españolas y 122 norteamericanas. El promedio total de problemas de accesibilidad detectados fueron de un 59% para las bibliotecas españolas y un 41% para las norteamericanas. (Caballero Cortés y otros, 2009).
- En 2013, sobre 16 sitios de universidades evaluadas, 7 españolas, 7 estadounidenses y 2 del Reino Unido, se determinó que los sitios cumplen entre el 25% y el 50% de los requisitos de accesibilidad. (Hilera y otros, 2014).
- Un estudio de 53 sitios de grandes hospitales españoles, (más de 500 camas), muestra que 27 no cumplen ningún nivel de adecuación. Solo 14 cumplen el primer nivel, 9 alcanzaron el primer y segundo nivel y 2 cumplen los tres niveles. (Calvo, 2014).
- Se realizó una investigación en 2010 sobre 102 sitios web orientados al turismo para personas que presentan algún tipo de discapacidad. Se estudiaron 40 cadenas hoteleras españolas, 49 hoteles de 4 o 5 estrellas y 13 compañías aéreas. Los resultados indicaron que el 65% de sitios no alcanza el nivel mínimo de accesibilidad. (Nadal y otros, 2011).
- En 2013, la Comisión Europea publicó un informe sobre un estudio realizado entre 2011 y 2013 de los sitios web de 31 países, 27 estados de la Unión Europea, Noruega, Australia, Canadá y Estados Unidos. Se incluyeron sitios de: recaudación de impuestos, servicios de empleo, bibliotecas, registros de automotores, becas estudiantiles, bancos, periódicos, ferrocarriles y trámites para documentación personal como pasaportes y cambios de residencia. Los resultados demostraron que: el porcentaje de accesibilidad de los países de la Unión Europea es de alrededor del 50%; el porcentaje de accesibilidad de los países no pertenecientes a la unión europea estuvo entre el 50% y el 60. (Kubitschke y otros, 2013)
- En la Universidad Nacional del Nordeste (Argentina) se evaluaron los sitios web municipales de la región NEA (Corrientes, Chaco, Misiones y Formosa). El cumplimiento oscila en un 6% en dos provincias, un 12% y un 18% respectivamente en las dos restantes. Los incumplimientos varían entre un 53% y 65%. (Fernández Vázquez y otros, 2012).
- En la Universidad Nacional de La Plata (Argentina), en el marco de un proyecto iniciado en 2007, entre la Facultad de Informática y la Comisión Universitaria de Discapacidad se analizaron los sitios pertenecientes a la UNLP. Sobre una muestra de 52 sitios, los resultados demostraron que solo un 21% resultaron accesibles. (Katz, y otros 2007). Otra investigación de la UNLP, publicada en 2008, evaluó la accesibilidad de los sitios web oficiales de 9 ministerios y 22 municipios de la Provincia de Buenos Aires. Los resultados reflejaron que, sólo 3 de los 31 evaluados resultaron accesibles. (Díaz y otros, 2008).
- Entre abril y junio del 2014, se realizó una evaluación del estado actual de 30 sitios de entidades bancarias

argentinas, públicas y privadas, para determinar el grado de adhesión a las normas recomendados en la Guía de Accesibilidad de Sitios Web del sector público nacional, versión 18.1. La evaluación se realizó con herramientas automáticas (TAW y HERA) y el análisis manual complementario. Se analizaron los puntos de verificación para los niveles de prioridad 1 y 2 de la página principal de cada uno de los sitios. Se concluyó que sólo el 24% de los sitios –2 públicos y 5 privados- se acercan a un nivel de adecuación A, por presentar un número de errores muy bajo en el nivel de prioridad 1. (Rossi, Chapetto, Curti, 2015).

Líneas de Investigación, Desarrollo e Innovación

La investigación de los portales se organiza en áreas de interés: instituciones bancarias, instituciones de salud, instituciones educativas, sitios gubernamentales, entre otros. Se determina el grupo a evaluar, se definen las muestras y se evalúan las aplicaciones informáticas a utilizar. Las herramientas de evaluación se utilizan para identificar problemas de accesibilidad en forma automática y se complementan con evaluación manual. Se definió trabajar inicialmente con TAW y Examinator.

El enfoque de la investigación es básicamente de tipo cuantitativo. Para la definición de los requisitos y criterios de cumplimiento se toma como base las Normas de Accesibilidad Web 2.0, recomendadas por la ONTI. El tipo de investigación es descriptiva. (Hernández Sampieri y otros, 2006).

Resultados y Objetivos

Como resultado se espera obtener un diagnóstico del estado de situación de los

sitios web argentinos, durante el período 2014-2018 de acuerdo con las recomendaciones realizadas en las Normas de Accesibilidad Web 2.0 propuestas por la ONTI

Hasta el momento, se ha avanzado con la evaluación de: bancos y redes sociales. Se prevé continuar con: universidades, empresas de servicios, empresas de turismo, noticias (diarios), empleos, hospitales.

Formación de Recursos Humanos

El proyecto está integrado actualmente por dos profesores adjuntos, cuatro docentes auxiliares, y alumnos. El proyecto tiene como objetivo la formación de recursos humanos a través de:

- la propuesta de temas y tutores para el desarrollo del Trabajo Final de Carrera de los alumnos de la Licenciatura en Sistemas de Información de la UNLu
- el desarrollo de temas de investigación para el desarrollo de tesis de maestría de los docentes y auxiliares que participan del proyecto.

Durante 2015, dos alumnos han comenzado a desarrollar su trabajo final de carrera y uno de los docentes se encuentra definiendo su propuesta de trabajo final de maestría.

Referencias

Argentina. Honorable Congreso de la Nación Argentina. Ley 26.653 de Accesibilidad de la Información en Páginas Web, 2010 (en línea). Argentina, Buenos Aires. Consulta: septiembre 2013.
<http://www.infoleg.gov.ar/infolegInternet/anexos/175000-179999/175694/norma.htm>

Argentina. Presidencia de la Nación Argentina. Acceso a la Información Pública. Reglamentación de la Ley 26.653 de Accesibilidad de la Información en Páginas Web, 2013, (en línea). Argentina, Buenos Aires. Consulta: septiembre 2013.
<http://infoleg.gov.ar/infolegInternet/anexos/210000-214999/210143/norma.htm>

Caballero Cortés, Laura; Faba Pérez, Cristina y Moya Anegón. Félix. Evaluación comparativa de la accesibilidad de los espacios web de las bibliotecas universitarias españolas y norteamericanas, 2009 (en línea). Investigación bibliotecológica vol.23, n.47, pp. 45-66. ISSN 0187-358X. UNAM, México. Consulta: 3/ 2014. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-358X2009000100003

Calvo-Calvo, M. A. (2014). Calidad y características de los sitios web de los hospitales españoles de gran tamaño. Revista Española de Documentación Científica, 37 (1): e032, doi. Consulta: marzo 2014 <http://dx.doi.org/10.3989/redc.2014.1.1049>.

Díaz, Javier; Harari, Ivana y Amadeo, Paola. Evaluación de Accesibilidad de Sitios Web Oficiales, Ministerios y Municipios de la Provincia de Buenos Aires, 2008, (en línea). LINTI, Laboratorio de Investigación en nuevas tecnologías informáticas. Facultad de Informática, Univ. Nac. de La Plata. Consulta: 3/2014. http://www.linti.unlp.edu.ar/uploads/docs/proceso_de_evaluacion_de_accesibilidad_de_sitios_web_oficiales.pdf

Fernández Vázquez, Agustina; Acevedo, Juan José; Mariño, Sonia; Godoy, María Viviana y Alfonso, Pedro. Comunicación y accesibilidad en sitios web municipales de la Región del Nordeste Argentino, su evaluación mediante validadores automáticos, 2012 (en línea). Questión. Revista especializada en Periodismo y Comunicación, Vol 1, No 35 (2012). Instituto de Investigaciones en Comunicación (IICOM). Facultad de Periodismo y Comunicación Social. Univ. Nac. de La Plata. Consulta: 3/ 2014. <http://perio.unlp.edu.ar/ojs/index.php/question/article/view/Article/1530>

Fundación CTIC. TAW. Servicios de accesibilidad y movilidad web. España. 2015. <http://www.tawdis.net/>

Hernández Sampieri, Roberto; Fernández Collado, Carlos y Baptista Lucio, Pilar. Metodología de la Investigación, 2006, 4ta. Edición. McGraw-Hill Interamericana. México. ISBN 970-10-5753-8

Hilera José R., Fernández, Luis, Suárez Esther y Vilar Elena T. Evaluación de la accesibilidad de páginas web de universidades españolas y extranjeras incluidas en rankings universitarios internacionales, 2013 (en línea). Revista Española de Documentación Científica, Vol.36, Nro.1. España. Consulta: marzo 2014. <http://redc.revistas.csic.es/index.php/redc/article/view/774/909>

Katz, S.; R. Vaena; I. Harari; D. Martorelli. Proyecto de accesibilidad web en la UNLP. 2007 (en línea). LINTI, Laboratorio de Investigación en nuevas tecnologías informáticas. Facultad de Informática, Universidad Nacional. de La Plata. Consulta: marzo 2014. http://www.linti.unlp.edu.ar/uploads/docs/proyectos_de_accesibilidad_en_la_unlp.pdf

Kubitschke, Lutz (empírica), Cullen, Kevin (WRC), Dolphin, Ciaran (WRC), Laurin, Susanna (Funka Nu) and Cederbom, Andreas (Funka Nu). Study on assessing and promoting e-accessibility. 2013 (en línea). Consulta abril 2014. <http://www.funkanu.com/PageFiles/17371/thestudyonassessingandpromotinge-accessibility.pdf>

Nadal, Gabriel Fontanet; Mayol, Jaime Jaume (España). Importancia y situación actual de la accesibilidad web para el turismo accesible. PASOS, [en línea]. 2011. Consulta: marzo 2014. http://www.pasosonline.org/Publicados/9211/PS0211_08.pdf.

ONTI. Oficina Nacional de Tecnologías de la información, 2011 (en línea). Jefatura de Gabinete de Ministros. Argentina, Bs As. Consulta: septiembre 2013 http://www.jgm.gov.ar/archivos/servicios_al_ciudadano/guia_accesibilidad_sitios_web_sector_publico_nacional_V18_1.pdf

ONTI. Oficina Nacional de Tecnologías de la información, 2014 (en línea). Jefatura de Gabinete de Ministros. Argentina, Buenos Aires. Consulta: 12/ 2015. <http://www.infoleg.gov.ar/infolegInternet/anexos/230000-234999/233667/norma.htm>

Rossi Bibiana, Chapetto Viviana, Curti Marcelo. “Accesibilidad de la información en los sitios web de entidades bancarias públicas y privadas de la República Argentina”. Universidad Nacional de Luján, Rosario, Argentina. JAIHO 2015 <http://44jaiio.sadio.org.ar/sites/default/files/sie34-48.pdf>

WAI. Web Accessibility Initiative. Web Content Accessibility Guidelines 1.0, 1999, (en línea). Massachusetts, USA. Consulta: septiembre 2013. <http://www.w3.org/TR/WAI-WEBCONTENT/>

WAI. Web Accessibility Initiative. WCAG 2 FAQ, 2013 (en línea). Massachusetts, USA. Consulta: septiembre 2013. <http://www.w3.org/WAI/WCAG20/wcag2faq#different>

W3C. World Wide Web Consortium. Introduction to Web Accessibility, 2005 (en línea). Editor: Shawn Lawton Henry, y participantes del Education and Outreach Working Group (EOWG). Consulta: diciembre 2013. <http://www.w3.org/WAI/intro/accessibility.php>

Un Método para la Evaluación de Modelos Conceptuales de Procesos de Negocio Basado en Lógica Difusa

Carlos Salgado, Mario Peralta, Daniel Riesco, Germán Montejano
Departamento de Informática Facultad de Ciencias Físico-Matemáticas y
Naturales Universidad Nacional de San Luis
Ejército de los Andes 950 – C.P. 5700 – San Luis – Argentina
e-mail: {csalgado, mperalta, driesco, gmonte}@unsl.edu.ar

Resumen

El modelado de procesos de negocio presenta una visión global de la organización que permite entender mejor la dinámica de la empresa y las relaciones que se dan internamente en la empresa y con su entorno. El modelado del negocio es la técnica por excelencia para alinear los desarrollos con las metas y objetivos de las organizaciones. Los modelos cumplen un rol fundamental en la especificación de los Procesos de Negocio (PN). Por ello, es de vital importancia la calidad de los mismos para que ayuden a mejorar el desempeño y evolución de la organización. Acorde a ello, se propone un método, basado en la Lógica Difusa, para evaluar modelos conceptuales de PN. Las distintas fases del método permiten hacer un control de la corrección de los modelos y dan soporte a los ingenieros y líderes de proyectos respecto de la conformación de los modelos de PN. Para ello, tiene en cuenta las buenas prácticas de modelado e intenta tener una aproximación al lenguaje humano para referirse a lo entendible, poco entendible o inentendible que puede ser un modelo. Ello posibilita la interacción de forma más comprensible entre los distintos miembros que intervienen en el modelado de los PN de una organización.

Palabras clave: Lógica Difusa, Procesos de Negocio, Modelado de Procesos de Negocio, Modelos Conceptuales.

Contexto

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto N° 22/F222. Dicho proyecto es la continuación de diferentes proyectos de investigación a través de los cuales se ha logrado un importante vínculo con distintas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el programa de Incentivos.

Introducción

El desarrollo de modelos conceptuales de PN constituye una de las tareas claves en las primeras etapas del ciclo de vida de los PN. Los modelos son utilizados como medios para que los participantes puedan entender fácilmente los procesos que representan. Además, son empleados como punto de partida a la hora de realizar cambios y adaptaciones de los PN a las nuevas necesidades de las empresas. Por ello, es un factor primordial que estos modelos sean de alta calidad, en cuanto a su entendibilidad y adaptabilidad.

Al hablar de calidad en el modelado conceptual, se debe distinguir entre la calidad del producto (relacionada con las características del modelo conceptual) y la

calidad del proceso (cómo se desarrollan los modelos) [1]. Moody en [2], establece que es fundamental que toda propuesta de evaluación adhiera a estándares aceptados y aplicados. Moody propone que deberían ser consistente con las normas de calidad ISO 9000 [3], e ISO/IEC 9126 [4], ya que un modelo conceptual es un tipo particular de producto (ISO 9000). En este trabajo se tuvo presente la norma actual de producto de software: ISO 25000. En base a ella, se tuvieron en cuenta las características externas, considerando de importancia la entendibilidad, la mantenibilidad y la modificabilidad de los modelos, e internas, principalmente acoplamiento, cohesión y tamaño de los modelos.

La complejidad de un modelo conceptual puede estar influenciada por los elementos que lo componen (tareas, subprocesos, participantes, eventos, etc.). Por lo tanto, no es aconsejable definir una medida general para su complejidad [5]. Rolón en [6] propone un conjunto de medidas para la calidad de modelos conceptuales de PN desarrollados en BPMN. Estas medidas se basan en la propuesta de García Rubio de medidas para la calidad de proceso software [7]. Desde este punto de vista, el tener métodos que permitan medir la calidad de dichos modelos será de gran ayuda para las organizaciones en cuanto a la administración, difusión y mantenimiento de los procesos que ellos representan. Desde esta perspectiva, el proceso de evaluación de requerimientos de calidad de los modelos conceptuales de PN es de suma importancia. Por lo tanto, será de gran utilidad contar con un método cuantitativo para la evaluación y comparación de las características deseables de todo modelo que se apoye en los principios y prácticas de la ingeniería de software.

En este sentido, en [8], presentamos un método para la evaluación de modelos de

PN en función de, independientemente de su representación. El método propuesto en [8] permite evaluar las principales características de calidad que se considera que todo modelo de PN debe satisfacer.

Sin embargo, y como argumentan Huang y Wu en [9], “la condición difusa de los pensamientos humanos puede ejercer cierta influencia en la respuesta de los expertos sobre sus preferencias con respecto a los factores, criterios y alternativas, y de sus cálculos y juicios subjetivos”. Respecto del modelado de PN, esta condición difusa del pensamiento humano, ejerce una influencia muy importante en cuanto a la percepción de las reglas del negocio, y por ende en la definición y construcción de los modelos que representan dichas reglas. Desde este punto de vista, el método no permite evaluar con certeza estas características difusas del razonamiento que pueden llevar a ambigüedades en los modelos.

La utilización de la Lógica Difusa en la evaluación de los modelos de PN, permitirá evaluar dichos modelos considerando dicha condición difusa, lo que permitirá dar una mejor valoración de la calidad de los modelos, y su aproximación a la realidad que representan, de manera que la interpretación de los modelos, y en consecuencia de la realidad que representan sea consistente y no se den interpretaciones ambiguas de ella.

Líneas de Investigación, Desarrollo e Innovación

En base a lo expresado previamente, en esta línea de investigación, se propone un método para evaluar modelos conceptuales de PN basado en la Lógica Difusa. El objetivo de la propuesta, es brindar un medio que ayude en la toma de decisión a la hora de evaluar la calidad de los modelos de PN. El método servirá

tanto para el análisis de los modelos de nuevos procesos, como para evaluar la adecuación de las modificaciones, ya sea debido a cambios en el proceso o a corrección de errores, que se puedan realizar a modelos existentes.

El método propuesto brinda ciertas ventajas al permitir hacer un control difuso. Un sistema de lógica difusa utiliza cualquier tipo de información y la procesa de manera similar que el pensamiento humano; por ello, los sistemas de lógica difusa son adecuados para tratar información cualitativa, inexacta e incierta, que permiten, además, tratar con procesos complejos, lo que la hace una alternativa interesante para modelar problemas de toma de decisiones.

El control difuso permite operar con conceptos vagos o ambiguos propios del razonamiento humano cualitativo, fundado sobre un soporte matemático que permite extraer conclusiones cuantitativas a partir de un conjunto de observaciones (premisas) y reglas cualitativas (base de conocimiento).

Cuando se cuenta con información imprecisa e insuficiente, usar instrumentos estadísticos no es suficiente para obtener resultados significativos. La Lógica Difusa surge precisamente para tratar con este tipo de problemas y lograr darles una solución óptima. De esta forma, una combinación entre un sistema de Lógica Difusa y la experiencia o conocimiento que tienen los encargados de tomar las decisiones es una excelente manera de obtener buenos resultados [10].

En el proceso de desarrollo de modelos de PN, muchas veces la información acerca de las reglas de negocio que se deben representar y modelar, suelen ser imprecisas o insuficientes, lo que lleva a modelos imprecisos. El uso de la Lógica Difusa en la evaluación de dichos modelos, permitirá, a través de los mecanismos que provee dicha lógica,

evaluar aquellas situaciones imprecisas, e incluso ambiguas, producidas en la construcción de los modelos.

Resultados Obtenidos y Objetivos

En base a lo expresado, se definió un método para el análisis y evaluación de los modelos conceptuales de PN basado en la Lógica Difusa.

El método propuesto se divide en 5 fases bien diferenciadas, comenzando con el establecimiento de los requerimientos de calidad a evaluar. Para hacer este estudio se tienen en cuenta las buenas prácticas de modelado de PN. Las buenas prácticas consideradas para el trabajo son tan amplias para modelado independiente del lenguaje, como también específicas a un lenguaje como BPMN.

En base a las buenas prácticas que surgen de estudios científicos, expertos del dominio y personal técnico de las organizaciones, se definen funciones de pertenencia que permitirán tomar decisiones en cuanto a la entendibilidad y mantenibilidad de los modelos de PN.

Para cada una de las variables que se desean estudiar en los modelos se define una función de pertenencia. Para definir dichas funciones se tuvo en cuenta que existen buenas prácticas generales que trascienden los lenguajes en las que sean aplicadas y que tienden a mejorar la comprensión de los modelos, así como a reducir los errores que se deriven del modelado. Por ejemplo, *7PMG: Seven Process Modeling Guidelines* [11], son guías de modelado definidas a partir de evidencia empírica en el modelado de procesos.

El uso del estándar BPMN proporciona a las organizaciones la capacidad de comprender y comunicar sus PN internos de manera estándar en una notación gráfica. Sin embargo, el uso de dicho estándar, no garantiza que los procesos se

modelen de forma clara y eficaz. La forma en que los modeladores interpretan las condiciones de negocio y cómo definen su estructura, es crucial para asegurar que se entienden correctamente. Por ello, y en base a estas guías y buenas prácticas, se definen las reglas borrosas que especifican el vínculo entre las variables de entrada y salida del sistema.

La interpretación de una regla si-entonces involucra dos pasos: (1) evaluar el antecedente mediante la aplicación de cualquier operador difuso y (2) implicar o aplicar el resultado del antecedente al consecuente. Esto se hace evaluando la función de pertenencia. Es decir, se trata de evaluar la activación de una regla en función del grado de cumplimiento del antecedente. Para realizar dicha tarea se usan operadores de composición de conjuntos difusos y se aplica un sistema de inferencia. En esta etapa del proceso, las salidas de cada una de las reglas se combinan para obtener un único conjunto difuso. Las entradas del proceso de agregación son las funciones de pertenencia truncadas obtenidas de la etapa de inferencia para cada una de las n-reglas.

Las operaciones básicas realizadas con conjuntos difusos se llevan a cabo mediante la aplicación de algún operador binario clasificado como T-normas (para operaciones de intersección) o S-normas (para operaciones de unión) [12, 13, 14, 15, 16]. En esta etapa se obtiene un valor nítido o concreto a partir del conjunto difuso de salida, el cual proporciona la solución del sistema planteado.

En la fase final del método, se debe realizar un análisis y comparación de los resultados obtenidos en la evaluación de los modelos respecto de las preferencias de los usuarios, obtenidas en la aplicación del método. Además, se debe documentar el proceso de evaluación y los resultados obtenidos, de manera que dicha

documentación sirva como referencia e historial de la evolución de los modelos de proceso de negocio estudiados en futuras evaluaciones de dichos modelos. Esta documentación puede servir como punto de referencia y comparación a la hora de evaluar nuevos modelos y procesos de negocio. Esta fase trata con actividades de análisis y comparación de las preferencias de calidad y los resultados obtenidos. A partir de las metas establecidas y el punto de vista de los interesados en los modelos y procesos de negocio a evaluar, esta etapa culmina con las conclusiones y recomendaciones del caso.

Esta etapa es una de las actividades más relevantes del método. Por ello, es de suma utilidad tener la información recopilada durante la aplicación del método volcada en estructuras y representaciones que sean claras de leer e interpretar. Desde esta perspectiva, se propone un formulario tipo que debería llenarse una vez realizada la evaluación de los modelos. Dicho formulario permite, entre otras cosas, tener presente qué funciones de pertenencia se utilizaron; si fueron definidas por el grupo evaluador o si se utilizaron otras definidas y almacenadas en un repositorio previamente. Además, se registran datos de los modelos, de los evaluadores, y si existen evaluaciones previas se incluye una referencia a ellas.

Con el objetivo de validar el método propuesto, se está trabajando en la aplicación del mismo en casos reales de la industria del medio. Además, en la continuidad de esta propuesta, se espera proseguir con el trabajo de mejora del método con la construcción de una herramienta que permita su aplicación automatizada en el análisis de los modelos estudiados.

Formación de Recursos Humanos

Bajo esta línea, en el grupo de

investigación se presentó un trabajo final de Especialización en Ingeniería de Software, [17], y una tesis de Maestría en ingeniería de Software referente al Modelado de Procesos de Negocio [18]. Actualmente, se está trabajando en una tesis doctoral en la que se combina el análisis de calidad de modelos de PN con la lógica difusa. Además, se está desarrollando una tesis de maestría en lo referente, como así también se están llevando a cabo algunas tesinas de grado para la Licenciatura en Ciencias de la Computación.

Referencias

- [1] M. Piattini, F. Ó. Garcia Rubio, and I. Caballero, *Calidad de Sistemas Informáticos: Alfaomega-RA-MA*, 2007.
- [2] D. Moody, "Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions". *Data & Knowledge Engineering*. Elsevier B.V., 2005.
- [3] ISO, "ISO Standard 9000-2000: Quality Management Systems: Fundamentals and Vocabulary, International Standards Organisation (ISO)." 2000.
- [4] ISO/IEC, "ISO/IEC Standard 9126: Software Product Quality, International Standards Organisation (ISO), International Electrotechnical Commission (IEC)," 2001.
- [5] Fenton, "Software Measurement: A Necessary Scientific Basis," *IEEE Transactions on Software Engineering*, 20(3), pp. 199-206, 1994.
- [6] E. Rolon, F. Ruiz, F. Ó. Garcia Rubio, and M. Piattini, "Aplicación de Métricas Software en la Evaluación de Modelos de Procesos de Negocio," *Revista Electrónica de la Sociedad Chilena de Ciencia de la Computación*, 2005.
- [7] F. Ó. García Rubio, "FMESP: Marco de Trabajo Integrado para el Modelado y la Medición de los Procesos Software," Doctoral, Departamento de Informática, Universidad de Castilla La Mancha, España, Ciudad Real. España, 2004.
- [8] N. Debnath, C. Salgado, M. Peralta, M. Berón, D. Riesco, and G. Montejano, "MEBPCM: A Method for Evaluating Business Process Conceptual Models. A Study Case". 9th ITNG, Las Vegas, Nevada, USA, 2012.
- [9] L.-C. Huang and R. Y.-H. Wu, "Applying fuzzy analytic hierarchy process in the managerial talent assessment model – an empirical study in Taiwan's semiconductor industry.," *International Journal of technology Management.*, vol. 30, pp. 105-130, 2005.
- [10] B. Kosko, "Pensamiento borroso: la nueva ciencia de la lógica borrosa," *Barcelona: Crítica.*, 1995.
- [11] J. Mendling, H. Reijers, and W. van der Aalst, "Seven Process Modeling Guidelines," *Information & Software Technology*, vol. 52(2), pp. 127, 2010.
- [12] A. Kaufmann and J. Gil Aluja, "Las matemáticas del azar y la incertidumbre: elementos básicos para su aplicación en economía," *Madrid: Ramón Areces*, 1990.
- [13] E. Trillas, "Conjuntos Borrosos," *Madrid: Vicens Vives.*, 1980.
- [14] J. Jang, E. Mizutani, and C. Sun, "Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence," *New York: Prentice Hall*, 1997.
- [15] A. Kulkarni, "Computer vision and fuzzyneural systems," *New York: Prentice Hall*, 2001.
- [16] N. Kasabov, "Foundations of neural networks, fuzzy systems and knowledge engineering," *Cambridge: The MIT Press*, 1998.
- [17] C. Salgado, "Tesis de Especialización en Ingeniería de Software: El Modelado de Procesos de Negocio: Aplicando LSP para la Evaluación de Lenguajes de Modelado de Procesos de Negocio," F.C.F.M.yN., UNSL, 2010.
- [18] C. Salgado, "Tesis de Maestría en Ingeniería de Software: MEMPN: Método para la Evaluación de Modelos Conceptuales de Procesos de Negocio ", F.C.F.M.yN., UNSL, San Luis, Argentina., 2013.

Estudio de los Modelos Conceptuales de Procesos Workflow para Analizar y Evaluar su Migración a la Nube

M. Peralta, C. Salgado, L. Baigorria, G. Montejano, D. Riesco
Departamento de Informática Facultad de Ciencias Físico-Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950 – C.P. 5700 – San Luis – Argentina
[mperalta, csalgado, flbaigor, gmonte, driesco]@unsl.edu.ar

Resumen

Dada la globalización de la información, las organizaciones tienden a virtualizar sus negocios: subir su negocio a la Nube. Desde la perspectiva de la complejidad de los Procesos de Negocio (PN), una de las tecnologías más significativas para soportar su automatización son los Sistemas de Gestión Workflow, dando soporte computacional para definir, sincronizar y ejecutar las actividades de un PN. Para favorecer y dar flexibilidad a dichos sistemas, es fundamental tener herramientas para medir su calidad. Por ello, proponemos un marco de trabajo para el modelado y medición de Procesos Workflow (PW) que ayude a la mejora y mantenimiento de los modelos y de los procesos que representan. En este marco, se define un conjunto de métricas para la medición individual de cada aspecto relevante para el modelado, como así también la complejidad estructural global del PW y los modelos que lo representan. Debido a que toda ejecución de un PN tiene un workflow subyacente, y al creciente trabajo en la nube, aplicamos nuestro framework para evaluar PW en la nube. Además de los procesos, se analizaron los servicios de cloud computing que los soportaban, debido a que su calidad está influenciada directamente por la calidad de dichos servicios.

Palabras clave: Workflow; Sistema de Gestión Workflow; Proceso de Negocio; Métricas; Cloud Computing; indicadores de rendimiento.

Contexto

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto N° 22/F222. Dicho proyecto es la continuación de diferentes proyectos de investigación a través de los cuales se ha logrado un importante vínculo con distintas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el programa de Incentivos.

Introducción

Analizando el ciclo de vida de los PN [1], es de gran importancia llevar a cabo una mejora continua de los mismos. Ello ha llevado a las organizaciones a buscar herramientas que proporcionen el soporte necesario para poder realizar dichas mejoras. Hoy en día, la Gestión de Procesos de Negocio (GPN) proporciona este soporte mediante los Sistemas de Gestión de Procesos de Negocio.

Una de las tecnologías más significativas para dar soporte a la GPN,

son los Sistemas de Gestión Workflow (SGW) que dan soporte a la automatización de los PN. Todo SGW debe garantizar que la organización realiza las tareas correctas, en el momento y de la forma adecuada. Por ello se considera fundamental tener una buena representación del proceso en la que se incluyan todos los aspectos de interés para el mismo. Para poder lograr dicha representación es necesario disponer de un lenguaje de modelado de procesos adecuado que, además, permita adaptar esa representación a los cambios continuos que los negocios de hoy experimentan.

Otro de los aspectos fundamentales en todo PW, es tener medios que permitan medir sus elementos más relevantes con el fin de detectar las áreas y aspectos a mejorar del proceso y, de esta manera, promover su mejora continua. Para ello, creemos que es necesario proveer un marco que permita realizar dicha medición. De acuerdo a esta necesidad, siguiendo la metodología propuesta en [2], hemos definido un conjunto de métricas elementales que sirven como indicadores de la complejidad estructural de los modelos de procesos workflow.

Desde otra perspectiva, debido a la globalización de la información y la comunicación, las empresas están dirigiendo su mirada a la administración de sus procesos en la nube. Desde el punto de vista de la administración de los procesos, un workflow puede ser controlado de forma manual, informatizada, o como una combinación de ambos métodos. Un esquema workflow permite a las empresas organizar las tareas y recursos a través de reglas que facilitan el control de los PN de la empresa. Es decir, que con un PW se logra un control total y absoluto de todas las tareas. En este sentido, las empresas interesadas en llevar su negocio a la nube, necesitan mantener ese control. Por ello, es fundamental tener PW de alta calidad.

Desde esta perspectiva, un medio para obtener PW de alta calidad, fácilmente mantenibles y adaptables, es proveer modelos de calidad de dichos procesos. Esto es válido para todo PW independientemente de dónde esté alojado. En particular, si los procesos serán administrados en la nube, será de gran utilidad tener herramientas que permitan evaluar la calidad de dichos procesos. Desde esta perspectiva, surge la pregunta: *es posible o no subir todo los procesos a la nube*, y en caso de no ser posible, se debe determinar *cuáles de esos procesos subir*.

Con la aplicación de las métricas se pretende mostrar y determinar en qué grado el modelado de PW ayuda en la mejora continua de los PN que se quieren llevar a la nube. Esto ayudará a proveer un medio para lograr procesos que sean más fácilmente mantenibles a partir de su entendibilidad y modificabilidad. Estas son propiedades muy importantes que deben tener los sistemas workflow actuales debido al dinamismo de los PN que ellos gestionan y que obligan al cambio y adaptación continua de estos sistemas. Más aún si se encuentran en la nube.

Desde otro punto de vista, en el ámbito de la medición, se pueden observar diversos trabajos en cuanto a la medición de los PN como en [3, 4]. Sin embargo, y a pesar de la importancia de medir la calidad de los PW, es muy poco el trabajo detectado en cuanto a la medición de calidad de estos procesos y de los modelos que los representan. Entre los trabajos en este campo se puede destacar las propuestas realizadas en [5, 6, 7], donde se propone una métrica para la medición de la complejidad del flujo de control basada en la complejidad ciclomática de McCabe. Otros trabajos que se pueden mencionar en este campo es la propuesta introducida en [8, 9]. En estos trabajos los autores introducen métricas para evaluar la cohesión interna de las actividades en un

PW y el acoplamiento entre sus actividades. Estas propuestas presentan alternativas para la medición de algunos aspectos del modelado de PW. En el ámbito de nuestro trabajo de investigación, hemos definido un conjunto de métricas para la medición de la calidad de modelos de PW desde el punto de vista de su mantenibilidad [10].

Respecto de Cloud Computing, en la actualidad existen diversas herramientas y frameworks que permiten trabajar los PW en la nube. En este campo, nuestra investigación radica en la definición de un framework que ayude a medir distintas características de los PW en la nube a través de la propuesta de un conjunto de métricas. El objetivo de dichas métricas es brindar una medición de la complejidad estructural de los PW y la relación de dichos procesos con su entorno. En este contexto, aplicamos las métricas propuestas en la evaluación de los PW de una empresa del medio, la cual pretende migrar a la nube sus procesos con el fin de incrementar su competitividad.

En todo PW, un aspecto fundamental es tener herramientas que permitan medir sus elementos más relevantes con el fin de detectar las áreas y aspectos a mejorar del proceso y, de esta manera, promover su mejora continua. Para ello creemos que es necesario proveer un marco que permita realizar dicha medición y tener criterios que ayuden a tomar esta decisión.

Líneas de Investigación, Desarrollo e Innovación

De acuerdo a las necesidades mencionadas previamente, siguiendo la metodología propuesta en [2], se definió un conjunto de métricas iniciales y elementales que servirán como indicadores de la complejidad estructural de los modelos de PW. Dichas métricas son una

adaptación de las propuestas en [11] para la medición de procesos software. Las métricas propuestas ayudarán a la evaluación, comparación y mejora de los modelos workflow y, en consecuencia, de los procesos que ellos representan.

En este contexto, hemos realizado experimentos aplicando dichas métricas para la evaluación y comparación de lenguajes de modelado [12, 13], como parte del proceso de validación práctica de las métricas propuestas. Además, se las aplicó en la evaluación de los modelos de proceso de una empresa del medio que, debido a las exigencias del mercado y a la fuerte tendencia de la globalización de los negocios, pretende subir sus PN a la nube.

Llevar el negocio a la nube brinda una nueva alternativa para poder ser competitivo en el mercado. En este contexto se adaptaron las métricas propuestas y se definieron nuevas métricas que permiten medir las características propias de los PN y aquellas inherentes a la nube, como por ejemplo, la comunicación entre procesos alojados en distintas nubes, la comunicación con otros procesos, entre otras.

Cloud Computing se puede pensar como un modelo de aprovisionamiento rápido de recursos de TI que potencia la prestación de servicios, negocio y TI, facilitando el trabajo del usuario final y del prestador del servicio.

Para lograr su objetivo, los proveedores de Cloud Computing proveen aplicaciones de negocio en línea que se acceden desde otro servicio Web o software como un navegador Web, mientras que el software y los datos se almacenan en los servidores. En este nuevo modelo de gestión de la información, los datos *sensibles* del negocio no residen en las instalaciones de las empresas, lo que podría generar un contexto de alta vulnerabilidad de la información. Desde el punto de vista de la GPN, las organizaciones empiezan a

adaptarse a esta nueva tendencia y necesitan “subir su negocio a la nube”. Sin embargo, debido a la estructura y las relaciones que se pueden dar en un PN, no siempre es posible llevar a la nube el negocio completo. Además, puede ocurrir que algunos procesos, que tienen algún grado de comunicación, estén implantados y administrados por distintos proveedores de cloud computing. En esta situación, es necesario determinar el grado de acoplamiento de las actividades alojadas en distintas nubes. Esto es importante desde distintas perspectivas, como el costo de comunicación, la seguridad y el resguardo de la información.

Bajo estas consideraciones, será de gran ayuda poder anticiparse a problemas o necesidades futuras. Para ello, es necesario complementar el estudio estructural de los PW con un análisis del rendimiento de esos procesos ejecutándose en la nube.

Para garantizar un buen rendimiento, es importante determinar la buena reputación comercial del proveedor como confiable, rápido, seguro y eficiente. Si falta alguna de estas métricas de rendimiento, el proveedor no tiene forma de chequear cuán bien se desempeña la aplicación en la nube. El mal rendimiento puede resultar en fallas inesperadas del servicio que dejan a los usuarios varados sin la información necesaria para la toma de decisiones. Cuando una de las métricas muestra signos de inclinación hacia resultados negativos, debería poder accederse a las herramientas para identificar los posibles problemas de la aplicación antes que los usuarios tengan acceso a ellas. Algunas métricas de rendimiento disponibles, en distintos materiales bibliográficos, incluyen las siguientes [14]: *Métricas de control de estado*, *Métrica de control de versiones*, *Umbral de recursos*, *El umbral del usuario*, *El umbral de solicitud de datos*, *El umbral de respuesta*.

Estas métricas se han incorporado al

marco de trabajo propuesto, para complementar el estudio de los modelos en las primeras fases de los proyectos, apoyando con información referente a los distintos servicios que brindan los proveedores donde serán luego implantados los PN. Esta información de performance sirve para hacer estimaciones, simulaciones y demás tareas para la toma de decisiones acerca de qué parte del negocio y en qué tiempo se traslada a la nube.

Resultados Obtenidos y Objetivos

En el contexto de nuestra investigación, consistente en definir un marco de trabajo para medir la complejidad estructural de los modelos de PW y la performance de los servicios prestados por los proveedores de cloud computing, proponemos un tablero de métricas para tal objetivo. Además, en el contexto de los PW en la nube, se definieron nuevas métricas y se utilizaron las propuestas en distintas fuentes bibliográficas que permiten medir el grado de acoplamiento y cohesión entre los procesos en la nube. Además, para las métricas de rendimiento se definieron indicadores que nos permitieron confrontar distintas evaluaciones de los servicios que brindan los diversos proveedores en los que se estaba interesado para migrar el negocio a la nube.

Este marco posibilita medir los PW en la nube. Se miden atributos tales como: cantidad de clientes accediendo, tiempos de respuesta, cantidad de recursos utilizados, acoplamiento, cohesión, cantidad de tareas / subprocessos por nube, etc. Todos estos valores permiten tener una medida de la calidad de los modelos, lo que posibilita trabajar en la mejora continua de los procesos que dichos modelos representan. También es necesario definir indicadores de calidad o benchmarking para poder comparar y mejorar cada uno de

los PN de las empresas u organizaciones. En este sentido, el análisis llevado a cabo permite tener información referente a la distribución de los procesos en la nube. Lo cual servirá de base para posteriores trabajos o mejoras de los mismos.

En la continuidad del trabajo, se aplicarán las métricas definidas a nuevos casos de estudio que lleven a una mejor validación práctica de las mismas. Además, se analizará la necesidad de definir nuevas métricas para evaluar otros aspectos de un PW, como por ejemplo la utilización y distribución de recursos.

Formación de Recursos Humanos

Basados en la temática planteada, se están desarrollando tesis de Maestría y Doctorado por parte de algunos integrantes del Proyecto. Se ha finalizado una tesis de Especialización en Ingeniería de Software [15]. En el marco de la Maestría en Ingeniería de Software que se dicta en la UNSL, dentro del contexto del Proyecto, se han dictado charlas destinadas a los maestrandos acerca de la temática de Modelado Workflow y sobre el trabajo que se está desarrollando al respecto. Además se están desarrollando tesinas de grado para la Licenciatura en Ciencias de la Computación.

Referencias

- [1] D. Georgakopoulos and A. Tsalgatidou, "Technology and Tools for Comprehensive Business Process Lifecycle Management". Workflow Management Systems and Interoperability. Springer V., 1998.
- [2] M. Serrano, M. Piattini, C. Calero, M. Genero, and D. Miranda, "Un método para la definición de métricas de software.," in 1er Workshop en Métodos de Investigación y Fundamentos filosóficos en Ingeniería del Software y Sistemas de Información, 2002.
- [3] E. Rolon, F. Ruiz, F. Ó. Garcia Rubio, and M. Piattini, "Aplicación de Métricas Software en la Evaluación de Modelos de Procesos de Negocio," Revista Electrónica de la Sociedad Chilena de Ciencia de la Computación, 2005.
- [4] E. Rolon, F. Ó. Garcia Rubio, F. Ruiz, and M. Piattini, "Validating a Set of Measures for Business Process Models Usability and Maintainability," 2006.
- [5] J. Cardoso, "How to Measure the Control-flow Complexity of Web Processes and Workflows," in Workflow Handbook 2005.
- [6] J. Cardoso, "Control-flow Complexity Measurement of Processes and Weyuker's Properties," 2005.
- [7] J. Cardoso, "Approaches to Compute Workflows Complexity". in Dagstuhl Seminar, The Role of Business Processes in Service Oriented Architectures, Dagstuhl, Germany, 2006.
- [8] H. A. Reijers, "A Cohesion Metric for the Definition of Activities in a Workflow Process.," Eighth International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design 2003.
- [9] H. A. Reijers and I. T. P. Vanderfeesten, "Cohesion and Coupling Metrics for Workflow Process Design," BPM 2004, LNCS 3080, pp. 290-305, 2004.
- [10] M. Peralta, F. Ó. Garcia Rubio, D. Riesco, C. Salgado, and G. Montejano, "Un Conjunto de Medidas para la Evaluación de Modelos Workflow," CACIC 2008.
- [11] F. Ó. García, "FMESP: Marco de Trabajo Integrado para el Modelado y la Medición de los Procesos Software". Departamento de Informática, Universidad de Castilla La Mancha. Ciudad Real. España, 2004.
- [12] M. Peralta, F. García, M. Piattini, and R. Uzal, "Un experimento Comparativo de dos Lenguajes de Modelado Workflow: YAWL vs Diagramas de Actividad," 8th ASSE, 2007.
- [13] N. Debnath, M. Peralta, C. Salgado, G. Montejano, D. Riesco, and M. Berón, "Metrics for Evaluation of Workflow Models: An Experiment for Validation," in 20th International Conference on Software Engineering and Data Engineering, Las Vegas - USA, 2011.
- [14] J. M. Myerson, "Best Practices to Develop SLAS for Cloud Computing.," IBM Systems Journal, 2013.
- [15] M. Peralta, "Los Procesos Workflow y su Modelado. Un Estudio de los Patrones Workflow en distintos Lenguajes de Modelado," Especialización en Ingeniería de Software, Departamento de Informática - F.C.F.M.yN., UNSL, San Luis, 2010.

Evaluación de la Usabilidad por medio de Usuarios Finales

Rocío A. Rodríguez¹, Pablo M. Vera¹, Isabel B. Marko¹,
Vicente R. Merchán Rodríguez², Gabriela Y. Valles¹

¹Grupo de Investigación GIDFIS, Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de La Matanza (UNLaM)
Florencio Varela 1903, San Justo, Buenos Aires, Argentina
{rrodriguez; pvera; imarko; gvalles} @ing.unlam.edu.ar

²Departamento de Ciencias de la Computación
Universidad de las Fuerzas Armadas (ESPE)
Av. Gral. Rumiñahui s/n Sangolquí, Quito, Ecuador
vrmerchan@espe.edu.ec

Resumen

Dado que cada vez es mayor el uso de dispositivos móviles, sobre todo en torno de la telefonía celular, es importante planificar soluciones acordes a las necesidades de los usuarios asegurando el acceso y usabilidad de las mismas. Evaluar la usabilidad es un proceso complejo el cual requiere de diversos pasos, los que se presentan en este artículo. Uno de dichos pasos será considerar a los usuarios finales de la solución y observar sus dificultades al interactuar con la misma. Motivo por el cual se plantean aspectos a considerar al momento de diseñar test de usabilidad que tengan un buen cubrimiento de los contenidos y permitan dejar en evidencia dichas dificultades.

Palabras clave: Usabilidad, Dispositivos Móviles, Aplicaciones Móviles, Sitios Web, Usuarios Finales

Contexto

El presente proyecto de Investigación y Desarrollo está radicado en la Universidad Nacional de La Matanza, llevado a cabo por el GIDFIS (Grupo de

Investigación y Desarrollo en Innovación de Software) perteneciente al DIIT (Departamento de Ingeniería e Investigaciones Tecnológicas). Estando asociado al equipo de trabajo un investigador externo, perteneciente a la Universidad de las Fuerzas Armadas de Ecuador.

Introducción

La usabilidad parte de un concepto más abstracto conocido como Uso. El Uso, en sus cuatro primeras acepciones, según el Diccionario de la Real Academia Española de la Lengua [1], es:

1. Acción y efecto de usar.
2. Ejercicio o práctica general de algo.
3. Modo determinado de obrar que tiene alguien o algo.
4. Empleo continuado y habitual de alguien o algo.

Es posible apreciar que todas las acepciones buscan determinar las acciones que permiten maniobrar algo, por supuesto en busca de un objetivo concreto. Particularmente la usabilidad identifica esas acciones que realizan las personas para utilizar un objeto, en esta

línea de I+D (Investigación y Desarrollo), ese objeto será un producto de software.

Definir a la usabilidad de manera concreta resulta complicado más aún cuando alrededor de este término existen algunos factores que inciden en este concepto. Lo cual ha llevado a una serie de escenarios de debate y discusión [2], [3], [4], [5]. Se destacan a continuación algunas definiciones:

- La usabilidad de un sistema o equipo es la capacidad en términos funcionales humanos para ser utilizado con facilidad y eficacia por un grupo específico de usuarios, dado determinado entrenamiento y apoyo a los mismos, para cumplir con un grupo específico de tareas, en un rango adecuado de escenarios ambientales [6].
- La ISO 9241-11 describe a la usabilidad como el grado con el que un producto puede ser usado por varios usuarios para alcanzar objetivos específicos con efectividad, eficiencia y satisfacción, en un contexto de uso específico [7].
- La usabilidad como el atributo de calidad que mide lo fáciles que son de usar las interfaces de las aplicaciones [8]. Siendo así, el nivel de usabilidad de un sistema es, por su parte, una medida empírica y relativa de la usabilidad del mismo, es decir, de la sencillez de uso y facilidad de aprendizaje para interactuar con éste [2].
- La ISO/IEC 25010 [9] que tiene un enfoque de calidad caracterizado por la calidad en uso al definirla como la capacidad del producto software para ser entendido, aprendido, usado y que resulta atractivo para el usuario, cuando

este se usa bajo determinadas condiciones [10].

Las aplicaciones móviles presentan a su vez mayores desafíos ya que los usuarios interactúan con ellas en contextos dinámicos, en muchos casos realizando otras tareas al mismo tiempo, con su atención no siempre centrada en el dispositivo. Lo cual indica que la forma de uso debe ser lo más simple y clara para que el usuario no se frustre al intentar utilizar el software en cuestión. Por otra parte tienen algunas limitaciones como por ejemplo el tamaño de pantalla y teclado reducido. En algunos casos el teclado no es físico sino virtual mostrándose en la pantalla. En las actuales pantallas táctiles los controles cercanos deben ser un elemento de análisis, ya que los usuarios deben con sus dedos lograr pulsarlos y la precisión en este caso es un problema.

Es importante poder asegurar "...el acceso a la información desde cualquier lugar, en cualquier momento e independientemente del dispositivo utilizado puede alcanzarse a través de aplicaciones que se adapten dinámicamente a las necesidades del usuario, a las capacidades del dispositivo y a las condiciones del entorno" [11].

Tomando en cuenta esto se deberán tener consideraciones especiales al momento de desarrollar una aplicación móvil. Por lo cual se establecen diversos pasos a llevar a cabo para poder testear la solución [12] Dichos pasos consisten en: (1) que el desarrollador podrá realizar una prueba sencilla en un browser de escritorio con soporte móvil, (2) existen validadores del W3C que permiten chequear el markup [13], Hojas de estilo (CSS) [14], link rotos [15], los cuales permitirán corregir errores existentes, (3) probar con equipos reales para saber cómo visualizará el usuario la solución en su dispositivo (en diversos tamaños de

pantalla, resolución...) y (4) probar con emuladores para ver cómo se comporta dicha aplicación en la mayor cantidad de dispositivos y versionados posibles. Por cada uno de los pasos anteriormente mencionados es probable que deban realizarse cambios para solucionar los errores encontrados. No obstante se incorpora un paso esencial (5) realizar pruebas con usuarios finales, esto permitirá conocer si la lógica aplicada al realizar una aplicación es comprensible por los potenciales usuarios de la misma. La estructura elegida para organizar las opciones de los menús, el nombre de las opciones, si los íconos son representativos, etc. (ver figura 1)



Figura 1. Proceso de desarrollo de una aplicación web móvil como una actividad cíclica.

Es por ello que es importante considerar la constitución de un laboratorio de Usabilidad, el cual permite que una persona proponga al usuario alcanzar un objetivo con un producto software y observar como lo hace (evidenciando las dificultades que se le presentan).

No es necesario contar con muchos medios para poder poner en marcha las

pruebas en un laboratorio de usabilidad, afirma Steve Krug [16].

Tomando esto en consideración se está trabajando en el diseño de test de usabilidad. El test de usabilidad deberá tomar en cuenta 3 aspectos:

1. Perfil de Usuarios: Pensar en las características de los potenciales usuarios, en este caso particular, que no necesariamente tengan conocimientos previos, en todas las franjas etarias.
2. Modalidad de las Preguntas: Se deben establecer preguntas para el usuario de distintos modos para que no sólo permitan que ellos accedan puntualmente a un contenido sino también que les propongan a los usuarios alcanzar un contenido para un uso específico.
3. Contenidos: El test de usabilidad deberá cubrir contenidos que estén en distinto nivel de profundidad, menues que tengan diferente cantidad de opciones y a su vez concentrarse en respuestas concretas que el usuario pueda obtenerlas leyendo el texto. Esto conlleva a tres características:
 - a. Distancia del Contenido: La cantidad de clicks para alcanzar un contenido puede alterar gravemente la Usabilidad de la aplicación. La regla establece una cantidad favorable de 3 click. De acuerdo a esta regla, los usuarios pueden dejar de navegar por el sitio web o utilizar la aplicación nativa móvil si su objetivo de búsqueda es de difícil acceso o si se encuentra a más de 3 clicks información.
 - b. Cantidad de ítems del menú: El cerebro humano tiene algunos límites en su capacidad de procesamiento de información. De acuerdo a los estudios realizados por George A. Miller en el año

1956, a corto plazo la memoria puede retener solamente entre 5 y 9 cosas por vez [17]. Es por ello que este argumento es usado para limitar el número de opciones navegables en los menús a 7 ± 2 ítems.

- c. Adecuación del Texto: Que tan rápido puede llegar el usuario a encontrar lo más importante del texto que está visualizado. La pirámide invertida es un modo de escritura cuando el resumen del artículo está presentado en el principio del artículo. Esto hace uso del “efecto cascada” (waterfall effect) donde se intenta dar a los usuarios una idea instantánea acerca de tema principal de la app. En el caso de una aplicación web, la página principal comienza con una conclusión, seguida de puntos clave y finalmente los detalles de menor importancia como información de fondo. Desde que los usuarios desean una instantánea satisfacción, la pirámide invertida respaldada por Jakob Nielsen, es muy importante para una mejor experiencia de usuario [18].

Todos estos conceptos permiten diseñar test que tengan el mejor cubrimiento de contenidos y puedan poner en evidencia las dificultades de los usuarios cuando interactúan con la solución.

Líneas de Investigación, Desarrollo e Innovación

Dentro de esta línea se plantean como temas de estudio:

- Usabilidad de soluciones móviles.
- Construcción de Métricas.

- Analizar Modelos de Usabilidad Existentes.
- Desarrollo una Metodología para la Evaluación de la Usabilidad.
- Estrategias para planificar Test con Usuarios Finales e Implementación de un laboratorio de Usabilidad

Resultados y Objetivos

Inicialmente el esfuerzo estuvo puesto en definir métricas que permitan evaluar las soluciones móviles. Esto consiste en generar documentación que indique de forma objetiva como dar respuesta a cada pregunta planteada asociada a la usabilidad de la solución. Claramente estas metodologías conllevan a la participación de evaluadores con el objetivo de analizar soluciones existentes. Pero el integrar a los potenciales usuarios y analizar las dificultades que ellos enfrentan al usar la solución, también da por resultado información que es vital para descubrir aspectos que atentan a la usabilidad y quedan por fuera del análisis que puede realizar un evaluador, tal como se expuso en la sección anterior.

Generar una metodología que permita diseñar test para soluciones existentes, mediante evaluación de las mismas por parte de usuarios finales es un gran desafío. Esto no es considerado como el único método de testeo de la solución sino como un paso que brindará información adicional sobre los problemas de usabilidad que presenta la solución.

Actualmente se está trabajando en el diseño de test de usabilidad, basados en soluciones tomadas como casos de prueba. Luego se podrá con esta experiencia elaborar un documento de estrategias para la conformación de los test de pruebas que serán aplicadas al momento de construir test para nuevas aplicaciones.

Formación de Recursos Humanos

El equipo de trabajo está constituido por 8 personas entre las cuales se encuentran alumnos de grado de Ingeniería con becas asignadas formándose en actividades de I+D. Por otra parte vinculado con esta temática se encuentran realizándose: 1 tesis Doctoral (UNLP) y 1 tesis Maestría (UAI).

Referencias

[1] RAE (2016), Diccionario de la Real Academia Española.
<http://lema.rae.es/drae/?val=uso>.

[2] Cuadrat Seix C. (2012), Estudio sobre Evaluación de la Usabilidad Móvil y Propuesta de un Método para Tests de Usabilidad Cuantitativos basado en Técnicas de Eyetracking, Universitat de Lleida.

[3] Enríquez, J. G., Casas, S. I. (2014). Usabilidad en Aplicaciones Móviles. Informes Científicos-Técnicos UNPA, 5(2), 25-47.

[4] Maniega-Legarda, D. Aplicación de criterios de usabilidad en sitios web: consejos y pautas para una correcta interpretación. Observatorio TIC: REBIUN Red de Bibliotecas Universitarias (2006).

[5] Mascheroni, M. A., Greiner, C. L., Petris, R. H., Dapozo, G. N., & Estayno, M. G. (2012). Calidad de software e ingeniería de usabilidad.

[6] B. Shackel y S. Richardson, Human factors for Informatics, Cambridge: Cambridge University Press, (1991),p. 24

[7] ISO, ISO 9241-11:1998(en), 1998.:
<https://www.iso.org/obp/ui/es/#iso:std:iso:9241:-11:ed-1:v1:en>

[8] Nielsen J. (1995), Ten Usability Heuristics for User Interface Design.

<http://www.nngroup.com/articles/ten-usability-heuristics/>

[9] ISO, ISO/IEC 25010:2011(en), 2011.
<https://www.iso.org/obp/ui/es/#iso:std:iso-iec:25010:ed-1:v1:en>

[10] M. Piattini, F. García, I. García y F. Pino (2012), Calidad de Sistemas de Información, Segunda ed. Alfaomega - RA-MA, Ed., México: Alfaomega grupo editor, S.A. de C. V. 2012, p. 372.

[11] W3C (2010), World Wide Web Consortium. Guía breve de web móvil.
<http://www.w3c.es/Divulgacion/GuiasBreves/WebMovil>

[12] Giulianelli, D. A., Pons, C., Vera, P. M., Rodríguez, R. A., Trigueros, A., Fernández, V., & Marko, I. (2012). Técnicas para la validación de un sitio Web Móvil. In XIV Workshop de Investigadores en Ciencias de la Computación.

[13] W3C (2013). “Markup Validation Service” v1.3+hg. <https://validator.w3.org/>

[14] W3C (2009). “CSS Validation Service” <http://www.css-validator.org/>

[15] W3C (2011). “Link CHEKER” version 4.81.
<https://validator.w3.org/checklink>

[16] Steve Krug, (2006). “No me hagas pensar. Una aproximación a la usabilidad en la Web”. Segunda Edición. Pearson Educación, Madrid.

[17] Miller George A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information
<http://cogprints.org/730/1/miller.html>

[18] Vitaly Friedman (2007) “30 Usability Issues To be Aware Of”
<http://www.smashingmagazine.com/2007/10/30-usability-issues-to-be-aware-of/> -

Mapas Conceptuales para tratar Nominalizaciones en Modelos de Requisitos

Graciela Hadad^{1,2}, Alberto Sebastián¹, Claudia Litvak^{1,2}, Jorge H. Doorn²

¹Facultad de Ingeniería y Tecnología Informática, Universidad de Belgrano
Federico Lacroze 1947 (C1426CPE), Ciudad de Buenos Aires, (011) 4511-4716

²Escuela de Informática, Universidad Nacional del Oeste
graciela.hadad@comunidad.ub.edu.ar, asebastian@outlook.com.ar,
claudia.litvak@comunidad.ub.edu.ar, jdoorn@exa.unicen.edu.ar

Resumen

La Ingeniería de Software ha experimentado una evolución en los procesos de desarrollo, focalizándose en campos como la Ingeniería de Requisitos. Uno de los pilares de este campo es la comunicación, por lo que se incentiva el uso de modelos escritos en lenguaje natural durante la definición de requisitos. Es conveniente entonces escribir estos modelos usando el propio lenguaje del universo de discurso bajo estudio. Si bien se logra así una mejora sustancial en la comunicación, surgen problemas inherentes al lenguaje natural, tales como las ambigüedades. Paralelamente, se ha comprobado, mediante estudios estadísticos de completitud, la ocurrencia de omisiones significativas en modelos del proceso de requisitos. Las actividades de verificación y validación suelen tener dificultades en la detección de estos problemas. Es por ello que se propone extender una técnica de inspección de modelos en lenguaje natural, basada en la construcción de mapas conceptuales, potencialmente útil para la detección de ciertas omisiones y ambigüedades provenientes de la nominalización de verbos, que se ha observado con

frecuencia en la comunicación y se traslada esencialmente a los modelos escritos en lenguaje natural.

Palabras clave: Ingeniería de Requisitos, Completitud de Modelos, Lenguaje Natural, Ambigüedad, Nominalizaciones.

Contexto

La propuesta que se presenta es parte de los proyectos de investigación “Estrategia para mejorar la completitud de modelos orientados al cliente en la ingeniería de requisitos” de la Universidad de Belgrano y “Tratamiento de los factores situacionales y la completitud en la ingeniería de requisitos” de la Universidad Nacional del Oeste.

Introducción

El proceso de definición de requisitos es considerado uno de los más complejos en el proceso de software [1], y es donde se presentan la mayor cantidad de defectos [2]. Este proceso se enfoca inicialmente en adquirir todo el conocimiento relevante del universo de discurso donde estará inmerso el

software, por lo que la comunicación es una actividad intensiva. Este se vale de diversas técnicas y modelos, dentro de los cuales, el Léxico Extendido del Lenguaje (LEL) permite mejorar la comunicación entre clientes e ingenieros de software [1]. Esto se debe a que se construye para representar, por un lado, y comprender, por otro, el vocabulario utilizado en el universo de discurso, y sirve de ancla para toda comunicación escrita (modelos e informes) y oral (entrevistas y exposiciones). Este modelo se suele construir al inicio del proceso de requisitos y, por lo tanto, su calidad tiene un impacto directo en los subsiguientes modelos que se construyan, como así también a lo largo de todo el proceso de desarrollo. Uno de los aspectos negativos en el proceso de requisitos son las omisiones y ambigüedades que se reflejan en los modelos [3, 4, 5] y que se manifiestan con mayor intensidad en aquellos modelos escritos en lenguaje natural [6].

En un proyecto de investigación de fines de los '90 se desarrolló un proceso de inspección para detectar defectos en el modelo LEL basado en formularios y guías de detección [7], el que se ha utilizado en cientos de casos con una tasa aceptable de detección de defectos. Sin embargo, en trabajos posteriores se puso en evidencia el bajo nivel de completitud que persistía en los LELs construidos [4, 5, 8, 9, 10, 11, 12], donde una buena parte de las omisiones correspondían a la categoría de ambigüedades. Debido a esto, se ha elaborado un proceso de inspección que se aboca principalmente a la detección de estos defectos, el cual está aún en etapa de prueba. No obstante lo cual, se ha observado a partir de resultados preliminares, que se requiere un estudio más profundo respecto a ciertos tipos de ambigüedades que suelen presentarse en estos modelos,

provenientes del uso de verbos y sus formas nominales (sustantivos) [13].

Líneas de Investigación, Desarrollo e Innovación

En el LEL se representan los términos relevantes utilizados en el universo de discurso, donde cada término se describe a través de los componentes: *nombre* formado por una palabra o frase (puede ser más de un nombre en caso de sinónimos), *noción* que describe el significado del término e *impacto* que describe cómo repercute en el universo de discurso. Estos términos se clasifican en 4 *tipos*: sujetos, objetos, verbos y estados.

A través de los trabajos realizados sobre completitud del modelo LEL mencionados anteriormente, se han observado algunas dificultades en el tratamiento de los verbos, ya que algunos ingenieros de software describen las actividades que se desarrollan en el universo de discurso utilizando la forma verbal y otros la forma nominal, lo que se denomina en lingüística *nominalización* del verbo [13]. Esta nominalización no depende estrictamente de aspectos cognitivos del ingeniero sino también del propio uso del término en el universo de discurso; cualquiera sea el caso la nominalización o no, puede provocar ambigüedades. La nominalización consiste en construir sustantivos a partir de verbos o de adjetivos [14, 15, 16, 17, 18]. Las diferencias de uso de estas derivaciones en los verbos pueden ocultar la existencia de resultados o efectos provenientes de dicha actividad, y/o pueden producir una mala interpretación en el uso nominal del verbo. En otras palabras, la nominalización de un verbo puede describir una *acción* o un *efecto* de dicha acción, y en ciertos casos esa nominalización puede involucrar más de

un significado: la *acción* y el *efecto*, es decir, se está en presencia de homónimos. Un ejemplo de nominalización del verbo “Facturar”, extraído de un LEL desarrollado para un instituto de salud, es el término “Facturación Mensual”, que podría interpretarse como la actividad de generar facturas o bien como el producto de esa actividad. Si este fenómeno es ignorado puede ocurrir y ocurre frecuentemente que quien construya el LEL sólo perciba uno de los significados posibles y en consecuencia omita impactos que luego redundarán en omisiones en otros modelos derivados de él y, finalmente, en omisiones en los requisitos del sistema. En la Figura 1, se presenta la definición de dicho término, donde en su contexto es utilizado como el efecto/resultado de la acción y, por ello, es clasificado como *objeto*. Asimismo, un verbo puede desencadenar en más de una nominalización. Por ejemplo: i) Envolver: Envoltura–Envoltorio; ii) Agrupar: Agrupación–Agrupamiento. Se debería evaluar qué nominalizaciones son usadas en un universo de discurso específico, a fin de evitar omisiones de términos en el LEL y el uso ambiguo de términos nominalizados.

<p>Nombre: FACTURACIÓN MENSUAL Tipo: Objeto Noción:</p> <ul style="list-style-type: none"> • Son los servicios prestados por el Instituto al prestador, en valores monetarios <p>Impacto:</p> <ul style="list-style-type: none"> • El Instituto la genera como resultado del proceso de facturar en el período de facturación correspondiente • Contiene el detalle de cada factura emitida
--

Figura 1. Término del LEL nominalizado

Por otro lado, en un proyecto de investigación en curso se ha propuesto un proceso de inspección del LEL basado en

la construcción de mapas conceptuales, que está fuertemente centrado en la detección de *omisiones* y *ambigüedades* [19]. Básicamente, consiste en construir un mapa conceptual [20, 21] de cada término definido en el LEL, donde las oraciones en cada definición son proposiciones en el mapa (ver ejemplo de la Figura 2 que representa el término descrito en la Figura 1). Luego, se sigue una secuencia de pasos para estudiar cada mapa y el conjunto de ellos, mediante un análisis visual, estadístico y semántico de conceptos que no son términos del LEL, del uso de conceptos que sí son términos del LEL y de las relaciones entre conceptos, a fin de identificar conceptos candidatos a términos del LEL, sinónimos no detectados, proposiciones ambiguas, omisiones menores de información y omisiones de términos candidatos del tipo verbo. Este proceso de inspección está en una etapa de prueba, llevándose a cabo un experimento controlado. No se ha incluido específicamente el tema de nominalizaciones en la inspección, por lo que se considera que sería conveniente estudiar la posibilidad de definir algunas guías de detección a través de los mapas conceptuales.

Resultados y Objetivos

Las consecuencias del uso no controlado de las nominalizaciones pueden producir un deterioro del proceso de requisitos, a través de una comunicación deficiente entre los clientes y los ingenieros.

El estudio de las nominalizaciones de verbos para ser representadas en el LEL ha llevado a establecer los siguientes casos:

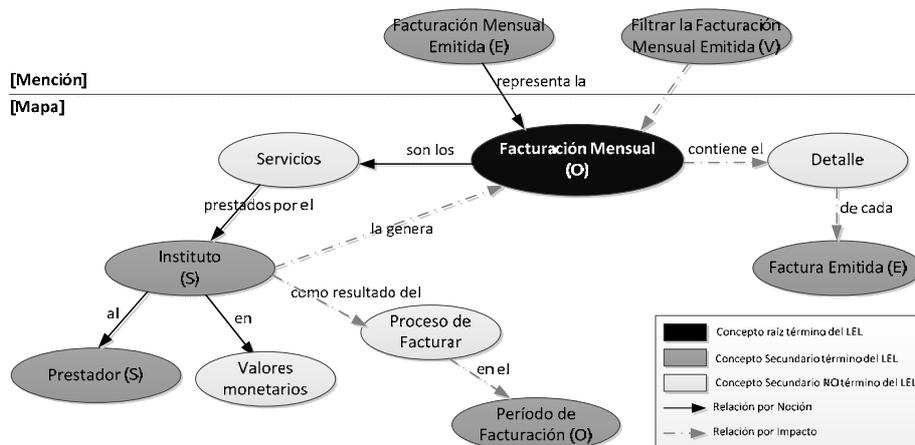


Figura 2. Mapa conceptual del término “Facturación Mensual”

- Se usa solo el verbo en el contexto: registrar el término tipo verbo en el LEL.
- Se usa solo la nominalización, siendo las posibilidades: a) representa solo la *acción*: registrar el término tipo verbo, b) representa solo el *efecto*: registrar el término tipo objeto o estado, o c) representa *acción y efecto*: registrar término tipo verbo y término tipo objeto o estado, siendo éstos homónimos.
- Se usa la forma verbal y la nominalización, siendo las posibilidades: a) representan solo la *acción*: registrar el término tipo verbo con el nombre de ambas formas (sinónimos), o b) representan *acción y efecto*: registrar ambas formas como término verbo y como término objeto o estado, considerando que la nominalización puede ser además un sinónimo de la forma verbal, produciendo entonces entradas al LEL homónimas.

Al elaborar el proceso de inspección del LEL usando mapas conceptuales se observó cierta dificultad en el tratamiento de conceptos y relaciones cuando el concepto era en sí mismo un verbo nominalizado. Esto ha llevado a la idea de realizar un estudio de las

nominalizaciones mediante mapas conceptuales, y cómo estos podrían ayudar a descubrir ambigüedades u omisiones de algunos términos relevantes en el universo de discurso debido a una interpretación deficiente de los verbos nominalizados o una visión parcial de los mismos. Las relaciones entre conceptos en un mapa conceptual aparentan ser más ricas de lo que a simple vista se aprecia [21] cuando los mapas se construyen a partir de un LEL. Esta riqueza podría inducir a descubrir principalmente verbos ignorados y, por otro lado, a cuestionar las nominalizaciones en los conceptos y su semántica en las proposiciones, de manera de llevar a capturar la existencia en el universo de discurso de su forma verbal o de otra semántica oculta.

Se considera que la detección de omisiones y ambigüedades a través de mapas conceptuales y la identificación de sus causas podrían derivar en nuevas heurísticas que colaboren en mejorar el nivel de completitud del modelo LEL.

Formación de Recursos Humanos

En el proyecto de la Universidad de Belgrano participan tres investigadores, uno de ellos en formación, mientras que el proyecto de la Universidad Nacional

del Oeste está integrado por cinco investigadores, uno de ellos en formación, y dos alumnos becarios.

La propuesta presentada es parte del Trabajo Final de Posgrado “Verificación de un modelo Léxico Extendido del Lenguaje con Mapas Conceptuales”, del Lic. Alberto Sebastián que está finalizando en la UCA, y de la Tesis de Doctorado “Gestión de la completitud en la Ingeniería de Requisitos”, de la Mg. Claudia Litvak que está desarrollando en la UNLP.

Referencias

- [1] Leite, J.C.S.P., Doorn, J.H., Kaplan, G.N., Hadad, G.D.S., Ridao, M.N. (2004). Defining System Context using Scenarios. En *Perspectives on Software Requirements*, Springer US, pp. 169-199.
- [2] Westland, J.C. (2002). The cost of errors in software development: evidence from industry. *Journal of Systems and Software*, vol.62, pp.1-9.
- [3] Firesmith, D. (2005). Are Your Requirements Complete? *Journal of Object Technology*, vol.4, nº1, pp.27-43.
- [4] Doorn, J.H., Ridao, M. (2003). Completitud de Glosarios: Un Estudio Experimental. 6th Workshop on Requirements Engineering Paracicaba, Brasil: Universidade Metodista de Piracicaba, pp. 317-328.
- [5] Hadad, G.D., Litvak, C., Doorn, J.H., Ridao, M. (2015). Dealing with Completeness in Requirements Engineering. *Encyclopedia of Information Science and Technology*, Third Edition. IGI Global, Mehdi Khosrow-Pour (ed), Information Science Reference, Hershey, pp.2854-2863.
- [6] Berry, D., Kamsties, E.(2004). Ambiguity in requirements specification. *Perspectives on software requirements*, Springer US, pp. 7-44.
- [7] Kaplan, G.N, Hadad, G., Doorn, J., Leite, J. (2000). Inspección del Léxico Extendido del Lenguaje. 3rd Workshop on Requirements Engineering, Río de Janeiro, pp.70-91.
- [8] Doorn, J.H., Ridao, M.N. (2009). Completeness Concerns in Requirement Engineering. *Encyclopedia of Information Science and Technology*, Second Edition. IGI Global, Mehdi Khosrow-Pour (ed.). Information Science Reference, pp.619-624.
- [9] Litvak, C.S., Hadad, G.D.S., Doorn, J.H. (2012). Un abordaje al problema de completitud en requisitos de software. XVIII Congreso Argentino de Ciencias de la Computación, Universidad Nacional del Sur, Bahía Blanca, pp. 827-836.
- [10] Litvak, C.S., Hadad, G.D.S., Doorn, J.H. (2013). Correcciones semánticas en métodos de estimación de completitud de modelos en lenguaje natural. 16th Workshop on Requirements Engineering. Universidad ORT, Montevideo, pp. 105-117.
- [11] Litvak, C.S., Hadad, G.D.S., Doorn, J.H. (2013). Mejoras semánticas para estimar la completitud de modelos en lenguaje natural. 1er CONAIIISI. Universidad Tecnológica Nacional, Córdoba.
- [12] Hadad, G.D.S., Litvak, C.S., Doorn, J.H. (2014). Problemas y Soluciones en la Completitud de Modelos en Lenguaje Natural. II Congreso Argentino de Ingeniería, Universidad Nacional de Tucumán, San Miguel de Tucumán, T366.
- [13] Real Academia Española y Asociación de Academias de la Lengua Española (2009). *Nueva gramática de la lengua española*, Madrid, Espasa Libros.
- [14] Rozwadowska, B. (1997). Towards a unified theory of nominalizations: External and internal eventualities (No. 2031). *Wdawn. Uniwersytetu Wrocawskiego*.
- [15] Alexiadou, A. (2001). *Functional Structure in Nominals. Nominalization and ergativity*. Amsterdam. John Benjamins Publishers.
- [16] Azpiazu, S. (2002). Las estrategias de nominalización y el adverbio: estudio contrastivo del caso español. Ediciones Universidad de Salamanca, España.
- [17] Bisetto A., Melloni C. (2005). Result Nominals: a Lexical-Semantic Investigation. Geert Booij et al. (eds.). *Fifth Mediterranean Morphology Meeting*, Fréjus, University of Bologna, pp. 393-412.
- [18] Rothmayr, A. (2009). *The structure of stative verbs (Vol.143)*. John Benjamins Publishing.
- [19] Sebastián, A., Hadad, G.D.S. (2015). Mejoras a un Modelo Léxico mediante Mapas Conceptuales. XXI CACIC, Universidad Nacional del Noroeste, Junín.
- [20] Novak, J., Gowin, D.B. (1988). *Aprendiendo a aprender*. Edic. Martínez Roca, Barcelona.
- [21] Novak, J., Cañas, A. (2008). *The Theory Underlying Concept Maps and How to Construct and Use Them*. Technical Report, Florida Institute for Human and Machine Cognition.

Técnicas de minería de procesos de negocio distribuidos con Bonita OS

Federico Gastón Madrid¹, Mg. Patricia Bazán², Lic. José Martínez Garro³

¹ ³ Facultad de Informática UNLP ² LINTI Facultad de Informática UNLP
fgmadrid89@gmail.com, pbaz@ada.info.unlp.edu.ar, josemartinezarro@gmail.com

Resumen

La madurez de los BPMS (*Business Process Management Systems*) ha hecho que las organizaciones se vuelquen cada vez con más confianza hacia modelos de gestión por procesos de negocio. La minería de procesos (*Process Mining*), una poderosa tecnología para administrar procesos operacionales no triviales, cuyo objetivo es extraer conocimiento de los registros de eventos obtenidos de distintos sistemas de información. Así, constituye un buen soporte para cerrar el ciclo de mejora continua que propone BPM.

Por otra parte *Cloud Computing* ofrece almacenamiento y cómputo distribuido bajo demanda, permitiendo descomponer un proceso y ejecutar cada una de sus partes de manera distribuida.

La minería de procesos distribuidos presenta un desafío importante de abordar y de validar este análisis aplicado a un BPMS concreto.

En este artículo se enuncia un enfoque para la aplicación de *Process Mining* sobre procesos distribuidos utilizando Bonita Open Solution como BPMS [1] [2] [3].

Palabras clave: *BPM, Bonita Open Solution, Web Services, Process Mining, Cloud Computing, XES, BI.*

Contexto

El presente trabajo se enmarca dentro de una línea de investigación que aborda características no clásicas de los procesos

de negocio como distribución de procesos y su disponibilidad en la “nube” y la socialización de procesos con el enriquecimiento de los rastros de ejecución. En este sentido, la minería de procesos aplicada a los procesos de negocio distribuidos, desplegados en un BPMS particular, es un caso de estudio que permite validar la línea de trabajo en la que se enmarca.

Introducción

El objetivo de *Business Intelligence* (BI) es el de generar información y conocimiento útil para el proceso de toma de decisiones complejas haciendo uso de los datos disponibles que deja el sistema a lo largo de su uso. Dentro de BPM el análisis de los rastros de ejecución conforman un punto clave dentro del ciclo de mejora continua. Técnicas de aplicación de BI son: BAM (Monitoreo de actividades de negocio), CPM (Gestión del rendimiento corporativo), CPI (Mejora continua de procesos), y BPI (Inteligencia de procesos de negocio) que permiten realizar reportes y tableros de mando. Luego para el análisis de cuestiones que son propias del proceso, los BPMS utilizan modelos de proceso para analizar los procesos operacionales. Pero estos modelos no son suficientes dado que se basan en un modelo idealizado y no en los datos de los eventos que ocurren con la ejecución de las instancias. Es en este escenario en el que *Process Mining* viene a dar una solución integral a la aplicación de

Business Intelligence a BPM combinando datos de eventos y modelos de proceso [3] [2] [4] [5].

La idea básica de Process Mining es extraer conocimiento del registro de eventos producidos por los sistemas de información, aplicando las técnicas:

- Técnicas de descubrimiento de proceso.
- Técnicas de chequeo de concordancia del proceso con la realidad.
- Técnicas de extensión y mejora del proceso.

Para lograr aplicar las distintas técnicas que nos brinda *Process Mining* debemos generar el registro de eventos en un formato específico que pueda ser procesado por una herramienta de minería de procesos. Conociendo la ubicación de los datos de interés, ya sea en la base de datos del BPMS (si ya contamos con un proceso desplegado) o la/s base/s de datos propia del sistema de información de la empresa, se genera el registro. En este punto es donde XES estandariza el modelo de dicho registro.

Por otro lado, la decisión de utilizar un BPMS en una organización para dar soporte a sus procesos de negocio puede ser arriesgada desde el punto de vista de la inversión que debe realizarse en software y hardware, así como en la capacitación de los recursos humanos para el despliegue y mantenimiento de dicho sistema. La escalabilidad también se ve comprometida, dado que el motor de procesos es capaz de atender simultáneamente una cantidad limitada de instancias de procesos [6].

Cloud Computing aporta una solución brindando servicios de computación bajo demanda con una alta fiabilidad, escalabilidad y disponibilidad en un entorno distribuido. El poder de cómputo, almacenamiento de datos y servicios se

contratan en una modalidad de “pago por uso”, evitando una gran inversión en software, hardware y mantenimiento, BPM basado en Cloud Computing (conocido como *Business Process as a Service – BPaaS*) es un modelo de servicio en el *cloud* donde las aplicaciones que se ofrecen son del tipo procesos de negocio o *workflows*.

Bonita Open Solution [10] es un BMPS de código abierto, para modelar procesos de negocio, exportarlos y ejecutarlos en un ambiente de producción. Esta herramienta se utilizará para aplicar los resultados de esta investigación. Por otro lado se eligió la herramienta ProM, también de código abierto para aplicar las técnicas de minería de procesos.

Para lograr combinar ambas tecnologías se plantea el siguiente escenario donde es posible realizar las siguientes tareas:

- Instanciar y monitorear procesos. Mediante un conector se instancia un proceso en un servidor remoto, permitiendo la ejecución de procesos distribuidos. Una aplicación PHP, recolecta la información de estos procesos y permite su monitoreo.
- Generar logs de eventos. Los procesos a medida que van ejecutando, mediante un conector, actualizan su propio log de eventos en el formato estándar XES.

En primer lugar un conector mediante el uso de la API REST del motor de Bonita [11], permite instanciar de manera remota un proceso en otro servidor. Luego en una base de datos local guarda las relaciones entre el identificador del proceso local con su par remoto para poder luego armar la cadena de monitoreo. Por otro lado otro conector mantiene un log por proceso en formato XES con los eventos ocurridos en las

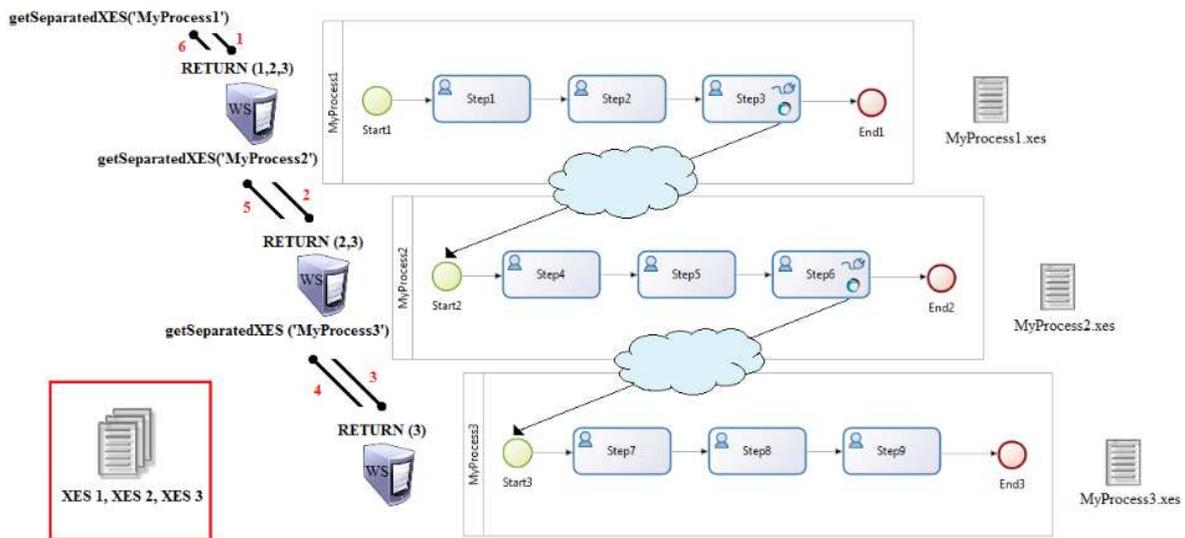


Figura 1: Arquitectura y comportamiento del WebService

instancias ya finalizadas listas para su procesamiento.

Dado este escenario se quiere extender la aplicación de monitoreo para que provea la funcionalidad de obtener el registro de eventos de un proceso distribuido. En este contexto se pueden observar dos posibles enfoques: análisis por partes y en conjunto. En el análisis por partes se obtendrían los registros de cada parte del proceso distribuido por separado (en archivos independientes) y cada uno de ellos sería una entrada diferente en la herramienta de minería de procesos, mientras que en el análisis en conjunto, mediante el procesamiento de cada registro, se produciría un solo archivo de eventos que representa el comportamiento del proceso completo, siendo este la única entrada a la que se le aplicaría la minería.

La primer alternativa brinda un análisis de alcance a nivel de servidor, es decir, sacar conclusiones sobre el rendimiento de esa parte del proceso en ese servidor específico en el que se encuentra (por ejemplo, si una tarea tarda mucho más de lo esperado en completarse puede deberse a que las prestaciones que da dicho servidor no son suficientes y por lo tanto

considerar el aumento de sus capacidades o mover esa tarea a otro servidor capaz de mejorar dicho rendimiento). También da lugar a evaluar aspectos del proceso como por ejemplo si la forma en que se decidió descomponerlo es la óptima para ese caso en particular o pueden hacerse algunos reajustes.

El segundo análisis es el más tradicional, dado que su resultado es el equivalente al del proceso unificado. En este caso se pueden observar aspectos tales como la detección de posibles cuellos de botella, rendimiento total, predicciones, etc.

Esta nueva funcionalidad se presenta como un servicio web cuyo comportamiento se muestra en la Figura 1. Utilizando el kit de herramientas NuSOAP [12] para desarrollar Web Services bajo PHP se ponen a disposición dos servicios: uno para obtener los registros de cada una de las partes en que está dividido el proceso y otro para obtener el registro del proceso unificado. En el primer caso basta con ir por cada servidor en donde se encuentra cada parte y recuperar el registro que actualiza el conector antes mencionado. En el otro caso, como XES tiene una estructura de

XML, se utiliza la clase SimpleXML [13] de PHP para el procesamiento de los registros existentes y generación del log unificado.

Líneas de Investigación, Desarrollo e Innovación

El presente trabajo continúa las líneas de investigación iniciadas en 2013-2014 en torno a Process Mining y BPM en el cloud [7] [8] [9]. Estas líneas de trabajo se han focalizado en: 1) la distribución de procesos, como mecanismo para mejorar la capacidad de escalar la complejidad de los mismos y también como la manera de utilizar y beneficiarse de nuevos enfoques de TI, como la computación en la nube, considerado como un nuevo modelo de servicios que afecta a toda la industria del software; y 2) la incorporación de aspectos colaborativos durante la ejecución de los procesos de negocios a fin de mejorar la participación de los actores y enriquecer los rastros de ejecución que retroalimentan el ciclo de mejora continua e, indirectamente, mejoran los modelos producidos.

Estos dos aspectos conducen a la definición de nuevos modelos y estrategias en torno a la gestión de procesos de negocio que requieren ser validadas sobre herramientas concretas. Tal es el caso de Bonita Open Solution.

Resultados y Objetivos

El resultado de este trabajo mejora la aplicación de minería de procesos en entornos distribuidos, favoreciendo su distribución en la nube. Asimismo es un aporte importante hacia la aplicación de técnicas de minería de procesos sobre los rastros de ejecución enriquecidos que producen los procesos de negocio colaborativos y sociales que son foco de la investigación actual.

En particular, este trabajo extiende el monitoreo de procesos a entornos distribuidos utilizando un BPMS de código abierto.

Formación de Recursos Humanos

BPM brinda un nuevo enfoque en la ciencia de los servicios de información, requiriendo una capacitación y formación de recursos humanos relacionados con esta tecnología. No sólo en aspectos de desarrollo de procesos BPM sino también la mejora continua de los mismos y la ampliación de sus ambientes de ejecución.

El presente trabajo se enmarca en una línea de investigación en Process Mining aplicado a procesos distribuidos donde se están formando alumnos para desarrollar su tesina e interactuar con investigadores formados con el objeto de incorporar herramientas de soporte de esta línea de trabajo para solucionar problemas reales.

Referencias

- [1] “Business Process Management: concepts, Languages, Architectures”. Mathias Weske Springer-Verlag Berlin Heidelberg 2007
- [2] “Process Mining Conformance and Extension”. O. by Anne Rozinat. - Eindhoven: Technische Universiteit Eindhoven, 2010. - Proefschrift.
- [3] “Process Mining Discovery, Conformance and Enhancement of Business Processes”. Wil M.P. van der Aalst Department Mathematics & Computer Science Eindhoven University of Technology Den Dolech 2 5612 AZ Eindhoven The Netherlands. 2011.

- [4] “Process Mining Project Methodology: Developing a General Approach to Apply Process Mining in Practice”. T.H.C. VAN DER HEIJDEN. BSc Industrial Engineering — TU/e 2011 Student identity number 0611037.
- [5] “Business Process Mining: From Theory to Practice”. Turner CJ, Tiwari A, Olaiya R, Xu Y. Business Process Management Journal 2012 Vol. 18 Iss:3, pp.493-512
- [6] T. Anstett, F. Leymann, R. Mietzner, and S. Strauch, “Towards bpm in the cloud: Exploiting different delivery models for the execution of business processes,” in Proceedings of the 2009 Congress on Services - I. Washington, DC, USA: IEEE Computer Society, 2009, pp. 670–677.
- [7] “Aplicación de técnicas de Process Mining para análisis de procesos de negocios desplegados en un BPMS”. Alumna Virginia Magliano, para su graduación como Licenciada en Sistemas de la Facultad de Informática de la UNLP. Octubre 2014.
- [8] “Técnicas de ejecución y monitoreo de procesos en Cloud BPM” [Martinez Garro J., Bazan P.] publicado en XVI Workshop de Investigadores en Ciencia de la Computación 2014: WICC 2014. Usuahia, Tierra del Fuego, Argentina. Mayo 2014. Con referato. ISBN 978-950-34-1084-4
- [9] “Ejecución y monitoreo de procesos de negocios distribuidos entre diferentes motores de Bonita OS” [Karabagosian L, Martinez Garro J., Bazan P.] publicado en XVI Workshop de Investigadores en Ciencia de la Computación 2014: WICC 2014. Usuahia, Tierra del Fuego, Argentina. Mayo 2014. Con referato. ISBN 978-950-34-1084-4
- [10] Documentación BonitaSoft - <http://documentation.bonitasoft.com/5x/index.php>
- [11] API REST de BonitaSoft - <http://documentation.bonitasoft.com/java/doc/rest/5.10/API/index.html>
- [12] NuSOAP - <http://nusoap.sourceforge.net/>
- [13] SimpleXML - <http://php.net/simplexml>

Evolución del software: La Accesibilidad Web en Sistemas Gestores de Contenidos de libre distribución

Sonia I. Mariño, María V. Godoy, Pedro L. Alfonzo

Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura.

9 de Julio 1449. CP: 3400. Corrientes. Argentina.

Universidad Nacional del Nordeste.

simarinio@yahoo.com, mvgg2001@yahoo.com, plalfonzo@hotmail.com

Resumen

Se presentan los avances en I+D+i centrados en la Accesibilidad Web, como paradigma de la sociedad de la información de interés multidisciplinar. Se fortalece la formación de recursos humanos en la temática, incorporando los estudios realizados en el desarrollo de becas otorgadas por la UNNE, tesinas de grado y asignaturas de la carrera.

Palabras clave: Accesibilidad Web, métodos y herramientas, formación de recursos humanos.

Contexto

En el marco de un proyecto de I+D+i acreditado por la Secretaría General de Ciencia y Técnica (UNNE), se indaga y aplican métodos y herramientas para medir y analizar la accesibilidad según pautas WCAG como un aspecto de la calidad de la Ingeniería del Software.

Introducción

La Accesibilidad Web (AW) referencia el acceso universal a este servicio de Internet, independientemente del tipo de hardware, software, infraestructura de red, idioma, cultura, localización geográfica y capacidades de los usuarios [1], [2].

En el marco del proyecto de “Sistemas de Información y TIC: métodos y

herramientas” se avanza en la indagación de métodos y herramientas y su aplicación con miras a aportar a la inclusión de los ciudadanos en el uso de herramientas informáticas en este siglo, y contribuir que las Tecnologías de la Información y Comunicación (TIC) estén al servicio de la comunidad para mejorar su calidad de vida.

Desde la perspectiva de la Ingeniería del Software (IS) es notable determinar la calidad de los productos software en proceso de elaboración, siendo la Accesibilidad Web una medida aplicable desde etapas tempranas de desarrollo y tratada como un requerimiento no funcional [3].

Otro aspecto de relevancia en la IS es el mantenimiento de sus productos para lograr mayor funcionalidad en la comunidad de usuarios.

Se considera al mantenimiento del software (MS) como “la modificación de un producto software después de la entrega para corregir fallos, para mejorar el rendimiento u otros [4]. Esta definición indica que las actividades de mantenimiento comienzan después que el producto está en funcionamiento. En ocasiones, algunas actividades pueden comenzar antes de la entrega del producto.

Los Sistemas Gestores de Contenidos o CMS son ampliamente aplicados en una diversidad de soluciones para mediar la comunicación con los usuarios finales.

En el mercado actual existen diversas herramientas comprendidas en esta categoría del software de libre distribución ampliamente utilizadas por la comunidad de programadores. Entre algunos CMS se mencionan Joomla [5], Drupal [6], Moodle [7], PhpFusion [8].

Además, se han reportado experiencias exitosas de su utilización en diversos dominios del conocimiento.

En este trabajo se considera relevante el estudio y difusión de contenidos referentes a la accesibilidad a los contenidos web; desde sistemas gestores de contenidos de libre distribución.

Cabe aclarar que la iniciativa de estudio de AW, desarrollada por este equipo de trabajo de la UNNE, coincide con el tratado en otras universidades ([9], [10], [11], [12]).

Líneas de Investigación, Desarrollo e Innovación

Para incrementar la generación e implementación de sistemas informáticos que respondan a estándares internacionales de accesibilidad como los establecidos por la WCAG 2.0 se:

- relevan, seleccionan y estudian métodos para el tratamiento de la AW basados en las pautas WCAG 2.0 [13].
- seleccionan, analizan y estudian herramientas informáticas utilizadas para la medición de accesibilidad, entre las que se mencionan: Brailersurf [14], Lynx Browser [15], TAW [16], HERA [17], EXAMINATOR [18], TextAloud [19], NVDA [20], WEBBIE [21], entre otras
- eligen y estudian diversas herramientas de programación y plataformas de libre distribución como Drupal [5], Joomla [6], Moodle

[7], ampliamente difundidas en el mercado para determinar su accesibilidad y promover mejoras.

- aplican las pautas WCAG 2.0 [13], en etapas tempranas del desarrollo de sitios web.
- utilizan diversos dispositivos móviles, teléfonos celulares, tablets que responden a diversas configuraciones para evaluar el nivel de accesibilidad de software basado en la web desde los mismos.

Resultados y Objetivos

La RedUNCI [22] ha incluido la Accesibilidad como un descriptor recomendado. El abordaje de la AW en el proyecto y su transferencia en espacios de educación superior -asignaturas Trabajo Final de Aplicación y Proyecto Final de Carrera- y con cursos de extensión, contribuye a su incorporación en la carrera y en la formación de recursos humanos.

Por otra parte dado que la asignatura Trabajo Final de Aplicación corresponde al plan de estudios anterior, se asegura que los alumnos de ambos planes de estudios accedan a contenidos y estrategias didácticas similares.

Los avances logrados en el año 2015, se plasman en producciones teóricas y experiencias prácticas.

Se avanzó en evaluaciones y definición de mejoras de CMS de libre distribución. Se han realizado presentaciones en congresos ([23], [24]) y en revistas ([25], [26], [27], [28] [29]) exponiendo los resultados.

Como se mencionó en trabajos previos, en el estudio y la evaluación de la accesibilidad se aplicaron medios automáticos y semi-automáticos y se procedió a la revisión manual en los casos

requeridos por la herramienta. Es decir, se aplicaron procedimientos híbridos de validación en los desarrollos web.

Formación de Recursos Humanos

Fomentar y desarrollar soluciones informáticas accesibles desde la Universidad, dirigidas a la comunidad, se considera relevante en el contexto de la sociedad del conocimiento. Se fortaleció la formación de recursos humanos en la temática, dado que se avanzó en:

- la revisión y profundización en métodos y herramientas de accesibilidad, y su introducción desde etapas tempranas del ciclo de vida de sistemas de información, actividad realizada con investigadores, becarios de grado y postgrado de la SGCyT – UNNE, y becaria CIN ([30], [31], [32], [33])
- en la asignatura Trabajo Final de Aplicación y Proyecto Final de Carrera, espacios donde se generan las tesinas para la obtención del título de grado, Licenciado en Sistemas de Información, se fomentó el abordaje de la temática considerando cómo un aspecto clave en soluciones web. Se enfatizó en su tratamiento como un aspecto integrado al modelo de ciclo de vida.
- se concretó la defensa de un Trabajo Final de Aplicación que abordó temas de investigación y aplicación de la AW [221].
- se avanzó en el desarrollo de un sistema para la gestión de AW en proyectos software [31].
- en una asignatura de programación del tercer año de la carrera para la elaboración de aplicaciones web se incorporaron aspectos relacionados con la accesibilidad desde etapas temprana del desarrollo. Se

mencionan el uso correcto de los códigos HTML y CSS. Se fomentó la comprobación del código de forma automática, a través de los servicios de validación suministrados por W3C.

Como líneas futuras de trabajo se mencionan el estudio de la Accesibilidad Web según las normas ISO, enfatizando la indagación de AW utilizando tecnología móvil.

Referencias

- [1] Consorcio World Wide Web (W3C). Disponible en: <http://www.w3c.es/>
- [2] Oficina Española. (2008). Word Wide Web - Guía Breve de Accesibilidad Web. Disponible en: <http://www.w3c.es/divulgacion/guiasbreves/accesibilidad>
- [3] Mariño, S. I., Godoy, M. V., Alfonzo, P., Acevedo, J., Gómez Solis, L., Fernández Vázquez, A. (2012) Accesibilidad en la definición de requerimientos no funcionales. Revisión de herramientas. *Multiciencias*, Vol. 12, Nº 3, 2012 (305 - 312)
- [4] IEEE STD 1219:1993. Standard for Software Maintenance. IEEE Computer Society Press. USA, 1993.
- [5] Drupal. Disponible en: <https://drupal.org/>.
- [6] Joomla. Disponible en: <http://www.joomla.org/>
- [7] Moodle. Disponible en: <https://moodle.org/?lang=es>
- [8] PhpFusion. Disponible en: <https://www.php-fusion.co.uk/home.php>
- [9] Russo, C., Sarobe, M., Saenz, M., Alonso, N., Pérez, Tessore, J., Cicerchia, B., Ado M, Ramon, H. (2014). Calidad, Usabilidad y

- Accesibilidad en los Entornos Virtuales de Enseñanza-Aprendizaje. Anales XVI Workshop de Investigadores en Ciencias de la Computación.
- [10] Rodríguez, N., Ale Levín, P. G., Jaen, I. (2014). Heurística para el desarrollo de Sitios Web Accesibles. Anales XVI Workshop de Investigadores en Ciencias de la Computación.
- [11] Hanari, I. (2011). Propuesta sobre aprender enseñando: desarrollo de un curso a distancia sobre Accesibilidad Web en manos de alumnos. Especialista en Docencia Universitaria, Facultad de Informática, Universidad Nacional de La Plata, Disponible en <http://sedici.unlp.edu.ar/handle/10915/19652>
- [12] Varas, V. D., Agüero, A. L., Guzmán, A. Elena, Martínez, M. (2015). Importancia y beneficios de la Accesibilidad Web para todos, X Congreso sobre Tecnología en Educación & Educación en Tecnología, p. 357-366,
- [13] Web Content Accessibility Guidelines (WCAG) 2.0. Disponible en: <http://www.w3.org/TR/WCAG20/>
- [14] Brailersurf. Disponible en: <http://www.snv.jussieu.fr/inova>
- [15] Lynx Browser. Disponible en: <http://invisibleisland.net/lynx/>
- [16] TAW. Disponible en: <http://www.tawdis.net/>
- [17] HERA. Disponible en: <http://www.sidar.org/hera/>
- [18] EXAMINATOR. Disponible en: <http://examinator.ws/>
- [19] TextAloud. Disponible en: http://www.nextup.com/files/manualtrans/Spanish_TextAloudManual.html
- [20] NVDA. Lector de pantalla. Disponible en: <http://nvda.softonic.com/descargar>
- [21] WEBBIE. Explorador web para usuarios ciegos y deficientes visuales. Disponible en: <http://www.webbie.org.uk/es/index.htm>
- [22] Red UNCI. (2014). Documento de Recomendaciones Curriculares de la RedUNCI 2014-2015. Disponible en: <http://redunci.info.unlp.edu.ar>,
- [23] Mariño, S. I., Godoy, M. V., Alfonzo, P. L. (2015). Accesibilidad Web. Aportes de las TIC al acceso universal. Primeras Jornadas de Intercambio de Experiencias en Inclusión y Abordaje de la Discapacidad, 2 y 3 de septiembre de 2015, Resistencia, Universidad Nacional del Nordeste
- [24] Mariño, S. I., Godoy, M. V., Alfonzo, P. L. (2015). Métodos y herramientas de las WCAG 2.0 en el desarrollo web. Workshop de Investigadores en Ciencia de la Computación, WICC 2015.
- [25] Casaro, D. E., Alfonzo, P. L. Mariño, S. I., Godoy, M. V. (2015). Applying WCAG 2.0 Guidelines in Online Banking Services-An Empirical Case Study in Argentina, International Journal of Information Science and Intelligent System, 4(3): 1-6, 2015
- [26] Casaro, D., Alfonzo, P., Mariño, S., Godoy, M. (2015). *Mantenimiento Correctivo Aplicado a un Sitio Basado en Joomla. Una Propuesta Centrada en la Accesibilidad*, Revista Latinoamericana de Ingeniería de Software, 3(2): 101-107

- [27] Duarte, P., Mariño, S. I., Alfonzo, P. L., Godoy, M. V. (2015). *Modelo de Proceso Software Aplicado a la Revisión de la Accesibilidad Web en desarrollos basados en IDE*. Revista Latinoamericana de Ingeniería de Software, 3(4): 177-182
- [28] Duarte, P., Mariño, S. I., Alfonzo, P. L., Godoy, M. V. (2015). Evaluación de la accesibilidad en software generado por un entorno de desarrollo integrado. *Técnica Administrativa*. Vol. 14, No. 1.
- [29] Duarte, P., Mariño, S. I., Alfonzo, P. L., Godoy, M. V. (2015). Automatic valuation of WCAG 2.0 guidelines in a Joomla FLOSS Platform. Inédito
- [30] Gomez Codutti A, Mariño, S. I, Alfonzo, P. (2015). Estudio de la Accesibilidad Web basado en criterios-pautas WCAG 2.0, Jornadas de Ciencia y Técnica de la Universidad Nacional del Nordeste, Beca de Pregrado.
- [31] Cavalieri, J. I., Alfonzo, P, Mariño, S. I. (2015). Sistema informático como soporte al análisis de las pautas de accesibilidad al contenido web. Beca de Pregrado otorgada por la Ciencia y Técnica de la Universidad Nacional del Nordeste.
- [32] Escalante, J. E., Mariño, S. I., Godoy, M. V. (2015). Valoración de accesibilidad del sistema SIGCES, utilizando pautas WCAG 2.0. Jornadas de Ciencia y Técnica de la Universidad Nacional del Nordeste. Beca de postgrado.
- [33] Duarte, P., Mariño, S. I., Alfonzo, P. (2014). TIC, GC y e-educación: Un modelo de sistema de información fundamentado en la Accesibilidad Web. Beca de estímulo a la investigación científica y desarrollo tecnológico de la Facultad de Ciencias Exactas y Naturales y Agrimensura – UNNE.

Business Intelligence & Knowledge Discovery en el contexto del estándar Project Management Body Of Knowledge del Project Management Institute

Moreno, Juan Pablo¹; Montejano, Germán²; Vilallonga, Gabriel^{1,2}

¹Departamento de Informática/ Facultad de Tecnología y Ciencias Aplicadas /
Universidad Nacional de Catamarca

Maximio Victoria 55 - San Fernando del Valle de Catamarca, +54 (0383) - 4435112

²Departamento de Informática/ Facultad de Ciencias Físico, Matemáticas y
Naturales / Universidad Nacional de San Luis

Ejercito de los Andes 950 - San Luis, +54 (0266) - 4520300

jpmoreno@tecno.unca.edu.ar; gmonte@unsl.edu.ar; gvilallo@tecno.unca.edu.ar/unsl.edu.ar

Resumen

Este trabajo se enfoca en el estudio de la experiencia previa y los elementos de conocimiento que existen en el Project Management Body of Knowledge (PMBOK) perteneciente al Project Management Institute (PMI).

Con esta investigación se pretende realizar aportes en el área de Gestión de Proyectos, planteando la necesidad de incorporar buenas prácticas en gestión del conocimiento y experiencias previas, que sirvan de soporte a una mejor gestión de proyectos de cualquier índole y envergadura. Para ello se propone un modelo de aplicación de técnicas y herramientas de Business Intelligence y Knowledge Discovery como buenas prácticas para gestionar la experiencia previa y el conocimiento de la organización en los procesos definidos en el Project Management Body of Knowledge. La adopción de este modelo repercutirá en una mayor precisión en los tiempos, costos y recursos que se asignan a un proyecto, mejorando las comunicaciones en el equipo de trabajo, mitigando riesgos y ofreciendo un mayor nivel de calidad del proyecto y del producto

Palabras clave: Gestión de Proyectos, Knowledge Discovery, Business Intelligence, Gestión del conocimiento

Contexto

El proyecto de investigación “Fortalecimiento de la Calidad y la Productividad en Ingeniería de Software”, posee una línea de investigación dedicada a los elementos de conocimiento que existen dentro de la Gestión de Proyectos. Dentro de ella es que se prevé trabajar sobre los estándares existentes en la Gestión de Proyectos.

El mencionado proyecto es evaluado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Catamarca y se desarrolla y ejecuta en ámbitos del Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa y en el Departamento de Informática de la Facultad de Ciencias Físico, Matemáticas y Naturales de la Universidad Nacional de San Luis y, donde los integrantes se desempeñan como docentes de la carrera de Ingeniería en Informática (UNCa), e Ingeniería en Informática y Licenciatura

en Ciencias de la Computación (UNSL) respectivamente.

Esta línea de investigación, inserta en el proyecto, establece como prioridad la formación científico-técnico de los integrantes con la premisa de proceder a la transferencia de resultados a los distintos actores sociales.

Introducción

Un proyecto se define como un esfuerzo para lograr un producto, servicio o resultado único en un tiempo establecido.

Existen varios estándares para la gestión y certificación de procesos:

- Capability Maturity Model Integration (CMMI)
- ISO 9001:2008
- Project Management Body of Knowledge (PMBOK)
- Project In Controlled Environments (PRINCE2)
- Rational Unified Process (RUP)
- Oracle Unified Method (OUM)
- Microsoft Solution Framework (MSF)
- Application Lifecycle Management (ALM)
- V-Modell

El hecho que PMBOK sea un estándar reconocido mundialmente (ISO21500), la flexibilidad de su implementación (adaptable desde un proyecto pequeño a uno complejo), la búsqueda de la calidad del proyecto y del producto que se obtiene como su resultado, y además su aplicación a cualquier tipo de proyecto, no solamente de software, motiva la selección del mismo para el desarrollo del trabajo.

En 1960 personas que realizaban gestión de proyectos en Estados Unidos, decide agruparse y conformar el Project Management Institute (PMI). Este grupo

empieza a documentar sus acciones y experiencias, confeccionando el PMBOK.

PMBOK es una guía donde se encuentran plasmadas las buenas prácticas concernientes a la Gestión de Proyectos (GP), las mismas surgen como resultado de recopilar experiencias de personas alrededor del mundo, que manejan proyectos de distinta índole. PMBOK está dividido en 10 Áreas de conocimiento y 5 Grupos de procesos, en total 47 procesos componen dichos grupos.

Áreas de conocimiento:

- Gestión de la Integración: incluye los procesos y actividades necesarios para identificar, definir, combinar, unificar y coordinar los diversos procesos y actividades de dirección del proyecto.
- Gestión del Alcance: incluye los procesos necesarios para garantizar que el proyecto incluya todo el trabajo requerido y únicamente el trabajo para completar el proyecto con éxito.
- Gestión del Tiempo: incluye los procesos requeridos para gestionar la terminación en plazo del proyecto.
- Gestión de los Costos: incluye los procesos relacionados con planificar, estimar, presupuestar, financiar, obtener financiamiento, gestionar y controlar los costos de modo que se complete el proyecto dentro del presupuesto aprobado.
- Gestión de la Calidad: incluye los procesos y actividades de la organización ejecutora que establecen las políticas de calidad, los objetivos y las responsabilidades de calidad para que el proyecto satisfaga las necesidades para las que fue acometido.
- Gestión de los Recursos Humanos: incluye los procesos que organizan, gestionan y conducen al equipo del proyecto.
- Gestión de las Comunicaciones: incluye los procesos requeridos para

asegurar la planificación, recopilación, creación, distribución, almacenamiento, recuperación, gestión, control, monitoreo y disposición final de la información del proyecto sean oportunos y adecuados.

- **Gestión de los Riesgos:** incluye los procesos para llevar a cabo la planificación de la gestión de riesgos, así como la identificación, análisis, planificación de respuesta y control de los riesgos de un proyecto.

- **Gestión de las Adquisiciones:** incluye los procesos necesarios para comprar o adquirir productos, servicios o resultados que es preciso obtener fuera del equipo del proyecto.

- **Gestión de los Interesados:** incluye los procesos para identificar a las personas, grupos u organizaciones que pueden afectar o ser afectados por el proyecto, para desarrollar estrategias de gestión adecuadas a fin de lograr la participación eficaz de los interesados en las decisiones y en la ejecución del proyecto.

Grupos de procesos:

- **Grupo de Procesos de Inicio:** Aquellos procesos realizados para definir un nuevo proyecto o nueva fase de un proyecto existente al obtener la autorización para iniciar el proyecto o fase.

- **Grupo de Procesos de Planificación:** Aquellos procesos requeridos para establecer el alcance del proyecto, refinar los objetivos y definir el curso de acción requerido para alcanzar los objetivos propuestos del proyecto.

- **Grupo de Procesos de Ejecución:** Aquellos procesos realizados para completar el trabajo definido en el plan para la dirección del proyecto a fin de satisfacer las especificaciones del mismo.

- **Grupo de Procesos de Monitoreo y Control:** Aquellos procesos requeridos para rastrear, revisar y regular el progreso y el desempeño del proyecto, para

identificar áreas en las que el plan requiera cambios y para iniciar los cambios correspondientes.

- **Grupo de Procesos de Cierre:** Aquellos procesos realizados para finalizar todas las actividades a través de todos los Grupos de Procesos, a fin de cerrar formalmente el proyecto o una fase del mismo.

Los grupos de procesos se ejecutan en distintas fases del proyecto y abarcan a todas las áreas de conocimiento descriptas anteriormente.

Los procesos poseen Entradas, Herramientas/Técnicas y Salidas, las entradas "Activos de los procesos de la organización" y "Factores ambientales de la empresa" en conjunto con la herramienta "Juicio de expertos", constituyen los elementos de conocimiento que posee la organización. Estos elementos de conocimiento se encuentran en 46 de los 47 procesos del PMBOK denotando la importancia de la experiencia y el conocimiento de la organización para poder llevar adelante un proyecto.

Debido a que los equipos de proyecto pueden variar de un proyecto a otro, se considera en el PMBOK una buena práctica la de conformar una Oficina de Gestión de Proyectos (PMO) dentro de la organización, que esté conformada por gerentes de proyectos de vasta experiencia, de tal forma que puedan llevar un control y seguimiento de los proyectos que la Organización se encuentra desarrollando. Esta oficina no se encontrará dentro del alcance del presente trabajo, debido a que la implementación de una PMO depende mucho del tamaño de la organización, no cualquier organización posee la cantidad de recursos y experiencia suficiente para su conformación, y además porque el estudio se centra en los procesos que componen el PMBOK. Sin embargo, el

resultado de esta línea de investigación podrá ser utilizado por la PMO ya que se pretende brindar herramientas que permitan la mejora del manejo de las experiencias previas en gestión de proyectos, función crucial de la PMO.

Líneas de Investigación, Desarrollo e Innovación

Si bien el PMBOK especifica o propone la forma en la cual se debe llevar adelante cada proceso, no especifica ninguna buena práctica para el manejo de los elementos de conocimiento enumerados anteriormente.

Una de las herramientas para la extracción de conocimiento es Business Intelligence (BI), el cual combina la recopilación de datos, almacenamiento de datos y la gestión del conocimiento con herramientas de análisis que permiten planificar y tomar decisiones. Estas decisiones están destinadas a, soportar las transacciones que realiza la Organización, tomar de decisiones y controlar la Organización e identificar nuevos negocios, no tan sólo para la Organización, sino también para aquellos que interactúan con ella.

Los grandes volúmenes de datos que la organización almacena son cruciales, y el tratamiento de los mismos para obtener información consume cada vez mayores recursos, por ello Knowledge Discovery (KD) es un proceso fundamental para la extracción de este conocimiento de forma automática. Este proceso engloba una serie de herramientas que permiten descubrir el conocimiento útil dentro de los datos.

Por lo anteriormente expuesto, se propone realizar un modelo de aplicación de Business Intelligence y Knowledge Discovery sobre los procesos del PMBOK, permitiendo incorporar el uso

de herramientas de BI y KD, que faciliten el uso y el aprovechamiento del conocimiento de la organización a fin de mejorar la gestión de proyectos con el estándar.

Resultados y Objetivos

El objetivo general de la línea de investigación inserta en el proyecto es proponer un modelo de aplicación de herramientas de BI y KD en los elementos de conocimiento que componen los procesos del PMBOK para una mayor precisión en los tiempos, costos y recursos que se asignan a un proyecto, mejorando las comunicaciones en el equipo de trabajo, mitigando riesgos y ofreciendo un mayor nivel de calidad del proyecto y del producto.

Objetivos específicos:

- Analizar y determinar las herramientas de BI y KD que sirvan para gestionar los elementos de conocimiento del proyecto.
- Desarrollar el modelo de aplicación de herramientas de BI y KD en los procesos del PMBOK.
- Validar el modelo de aplicación.
- Propulsar el interés por el tema planteado, organizar su desarrollo y sentar las bases para nuevas investigaciones.

Al finalizar esta investigación se pretende obtener un modelo que permita ajustar de mejor manera la forma en que se planifican los proyectos, utilizando toda la experiencia posible a través de las herramientas de BI y KD. Esto repercutirá en una mayor precisión en los tiempos, costos y recursos que se asignan a un proyecto, mejorando las comunicaciones en el equipo de trabajo, mitigando riesgos

y ofreciendo un mayor nivel de calidad del proyecto y del producto.

Formación de Recursos Humanos

El proyecto cuenta con integrantes docentes que se encuentran en la etapa de desarrollo de sus tesis de posgrado en el marco de la carrera de maestría en Ingeniería del Software en temas específicamente relacionados al área del proyecto. Además los participantes pertenecen a distintas cátedras de las carreras de ingeniería y licenciatura de las universidades que participan.

Se ha procedido a la incorporación de alumnos de los últimos años con la finalidad de incorporarlos a actividades de investigación y desarrollo en las áreas específicas del proyecto.

Específicamente el proyecto de investigación prevé el programa de capacitación y formación de recursos humanos, que contempla las siguientes actividades:

- Incorporación de alumnos de los últimos años de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa., en calidad de auxiliares de investigación.
- Dirección de tesinas de grado de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa integrados al proyecto.
- Participación de los integrantes del proyecto en cursos de actualización y posgrado en el área de estudio.

Para garantizar la capacitación y actualización del equipo de investigación, así como la difusión de los avances y resultados logrados, se propuso la participación en eventos nacionales e

internacionales de la especialidad, como congresos, simposios, seminarios y cursos.

Referencias

Project Management Body of Knowledge® - Fifth Edition – Project Management Institute - 2012

CMMI for Development: Guidelines for Process Integration and Product Improvement -Third Edition - Addison-Wesley Professional - 2011

CMMI for Acquisition: Guidelines for Improving the Acquisition of Products and Services, Second Edition - Addison-Wesley Professional - 2011

CMMI for Services: Guidelines for Superior Service, Second Edition - Addison-Wesley Professional - 2011

Managing Successful Projects with PRINCE2® - TSO (THE STATIONERY OFFICE) - 2009

The Rational Unified Process: An Introduction - Third Edition - Addison-Wesley Professional - 2003

Business Intelligence, A managerial approach - Second Edition - Prentice Hall - 2010

Business Intelligence, Competir con información - ESADE Business School - 2012

Knowledge Discovery Process and Methods to Enhance Organizational Performance - CRC Press - 2015

Principles of Data Mining and Knowledge Discovery - Springer Berlin Heidelberg - 2001

Conocimiento en Acción, Cómo las Organizaciones manejan lo que saben - Prentice Hall -2001

La organización creadora de conocimiento: Como las compañías japonesas crean la dinámica de la innovación - Editorial Oxford - 2009

La Dirección De Proyectos En Las Organizaciones - GRANICA - 2005

Definición de Métricas de Calidad para Productos de Software

Alejandra Velia LOPEZ¹, Alberto SÁNCHEZ², Germán Antonio MONTEJANO^{2,3}

1) Facultad de Tecnología y Ciencias Aplicadas, Universidad Nacional de Catamarca, Maximio Victoria 35, (4700), Catamarca, Argentina.

Tel: +54(3834)435112

e-mail: velialopez@gmail.com

2) Universidad Nacional de San Luis, Ejercito de los Andes 950, (5700) – San Luis - Argentina. Tel: +54(266)4520300 int. 2128

e-mail: alfanego@unsl.edu.ar - gmonte@unsl.edu.ar – web:

<http://www.unsl.edu.ar>

3) Universidad Nacional de La Pampa, Av. Uruguay 151, (6300) – Santa Rosa – La Pampa - Argentina. Tel: +54(2954)245220 int. 7125

Resumen

Desde la década del 90, las métricas del software y el proceso de medición asociado, han captado la atención de la comunidad de la Ingeniería de Software como medio para cuantificar y controlar la calidad del software [FN99, IEEE05, Pre05]. Pero para usar las métricas adecuadamente, no es suficiente con medir los atributos cuantitativamente, sino que es necesario tener en cuenta consideraciones como; unidades que se aplican, el tipo de software al que es aplicable, condiciones en que se deben recoger los datos, restricciones que el producto debe cumplir, etc.

Con esta investigación se pretende realizar aportes al área de calidad, a través de la formulación de métricas para productos de softwares

Palabras clave: Software, Métricas, Calidad.

Contexto

La calidad en la industria ha adquirido especial estatus y una relevancia que si bien no es inédita, adquiere nuevos matices en el variado mundo del software, como resultado de la competitividad y de la globalización.

Por otra parte en nuestro país esta industria se encuentra en pleno auge, con una fuerte demandada y en constante crecimiento, donde existen empresas de diferentes portes hasta pequeños emprendimientos con grandes oportunidades a nivel regional y exterior, pero cuyas puertas a la exportación están abiertas para aquellas que cuentan con alguna certificación que asegure la calidad de sus productos. Sin embargo puede ser un tema complejo si se tiene en cuenta la envergadura de la empresa tanto en recursos humanos como económicos,

además de otros factores como carga de trabajo y entorno o zona geográfica. Por lo que la calidad debe ser estudiada e incorporada desde los centros de generación de conocimientos como son las universidades, para que el proceso hacia la mejora continua sea un proceso natural de las industrias de software.

Esta investigación se enmarca en este contexto, en la definición de nuevas métricas para un sector específico como es el farmacéutico y busca incentivar la discusión académica y la investigación aplicada de temas relacionados a la calidad del software, en colaboración con la empresa V&S Ingeniería En Sistemas SRL que proveerá el sistema para las pruebas pertinentes, con el objeto de demostrar una investigación aplicada sacada de la problemática real del cual ha surgido esta investigación.

Esta línea de investigación se encuentra inserta en los *Proyectos de Ingeniería de Software de la Universidad Nacional de San Luis (UNSL)* y en el proyecto *Fortalecimiento de la Calidad y la Productividad en Ingeniería de Software* subvencionado por la Secretaría de Ciencia y Tecnología (SCyT) de la Universidad Nacional de Catamarca (UNCa), Argentina.

Introducción

En la industria de software, el mercado es cada vez más exigente y selecto, no sólo en lo que se refiere al precio, sino a la calidad que abarca importantes características, entre las cuales aportan a

la confiabilidad para adquirir productos de software y servicios posventa.

La implementación de un modelo o estándar de calidad implica costos, tiempos y recursos insumidos, pero las ventajas de implantar el modelo correcto supera el esfuerzo inicial, logrando productos de software y servicios con valor agregado, posicionando a la empresa en un mercado más competitivo, incrementando la productividad, mejorando la efectividad, reduciendo costos y asegurando la satisfacción de los clientes internos y externos, entre otros.

Sin embargo, la incorrecta elección de un estándar o modelo puede provocar serias consecuencias, inclusive el efecto contrario. Para evitar esto, previamente debe existir un estudio para la elección del modelo o estándar que se ajuste a las características, necesidades y objetivos de la empresa.

A lo largo del tiempo se han desarrollado varios modelos y estándares de Calidad, entre los modelos y estándares mas nombrados a nivel proceso se encuentra el Capability Maturity Model Integration (CMMi) que ha evolucionado hasta su versión actual V1.3, Personal Software Process (PSP), Team Software Process (TSP), Six Sigma For Software, TICKIT entre otros. Los estándares para el proceso en la familia de las ISO el 90003:2004, últimamente el revisado y modificado estándar 9001 en su versión ISO/IEC 9001:2015, ISO/IEC 12207, ISO / IEC TR 15504 y la ISO/IEC 20000:2005

Entre los modelos y estándares para el producto puede mencionarse el de McCall, Evans y Marciniak, Deutch y Willis, Furps, Boehm, entre otros

[PGC07]. Un hito en la definición de estándares de calidad del producto software, lo constituye la publicación del ISO 9126 en el año 1991. Luego, en el año 2001, este estándar fue reemplazado por dos estándares relacionados: el ISO/IEC 9126, que especifica características y métricas de la calidad del producto software; y el estándar ISO/IEC 14598, que define el proceso de evaluación de productos software.

El estándar ISO/IEC 9126 se compone de cuatro partes: modelo de calidad, métricas externas, métricas internas y métricas para la calidad en uso. Propone un modelo de calidad categorizando la calidad de los atributos software en seis características (funcionalidad, fiabilidad, usabilidad, eficiencia, mantenibilidad y portabilidad), las cuales son subdivididas en subcaracterísticas. El modelo más actual está representado por las normas ISO 25000, conocidas con el nombre de SQuARE (*Software Quality Requirements and Evaluation*), basada en ISO 9126 y en ISO 14598, se desagrega en 5 tópicos: 1-Gestión de la Calidad (2500n), 2-Modelo de Calidad (2501n), 3- Medidas de Calidad (2502n), 4- Requerimientos de Calidad (2503n) y 5-Evaluación de la Calidad (2504n) [PGC07].

Los modelos y estándares brindan la posibilidad de transformar la calidad en algo concreto, tangible, objetivo (por contraposición de subjetivo), que se puede definir, que se puede medir y que se puede planificar. Dan pautas para implementar programas de mejoras en búsqueda de la calidad a través de

procesos que serán definidos por la organización.

La necesidad de medir la fundamenta DeMarco [DeM82] en que “No se puede controlar lo que no se puede medir”. Y Fenton [FP97], nombra cuatro razones para medir procesos software, productos o recursos:

- Medir para caracterizar
- Medir para evaluar
- Medir para predecir
- Medir para mejorar.

Segun Sommerville [Som05], la medición es una actividad que forma parte de un proceso (ver Fig.1), que consiste en asociar valores numéricos a atributos de productos o procesos de software.

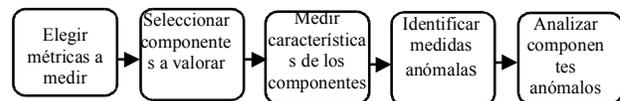


Figura 1: Proceso de medición definido por Sommerville

El glosario IEEE de términos de ingeniería del software define la ingeniería del software como la *aplicación de una aproximación sistemática, disciplinada y cuantificable al desarrollo del software...*. Lo que significa que la medición debe estar ligada a nuestra disciplina como una actividad necesaria a lo largo de todo el ciclo de vida del software. Y para concluir se puede decir que la medición de los productos y sus características hace posible la mejora de su calidad.

Líneas de Investigación, Desarrollo e Innovación

Existen numerosas publicaciones, artículos y conferencias sobre métricas del software donde se abordan los aspectos técnicos de la medición; pero no existen guías claras para la implementación de programas de métricas en las empresas. Definir métricas que se quieran implementar en un proceso de medición para conseguir medir la calidad de un producto o proceso de software, no es algo trivial. La línea principal de investigación de este trabajo propone formular métricas para un caso de estudio particular como es el dominio del mercado farmacéutico, para ser tomadas como referencia tanto para la calidad desde el punto de vista de la empresa informática que provee el producto, como para los clientes que necesiten adquirir un producto software de este dominio. Puntualmente para la comercialización de productos farmacéuticos, el cual cuenta con particularidades y características que difieren de otros modelos de negocios al tener que gestionar obras sociales, ART, convenios, trazabilidad, etc. La JARPHA (Journal of the American Pharmacists Association) resalta el rol de las farmacias como eslabón partícipe en la atención de la salud, por el valor de los servicios que ofrecen todos los días a millones de pacientes durante el proceso de dispensación, siendo el software parte vital de ese proceso de dispensación

Otra línea de investigación que surge primeramente es el estudio de los modelos y estándares de calidad

anteriormente nombrados, y una vez considerados se seguirán recomendaciones que sugieren algunos autores como [Fenton, Piattini, L. Briand, S. Morasca, etc.] y que se encuentran bien sintetizadas en [DDM12] cuando se definen métricas como:

- Las métricas se deben definir en función de un objetivo claro.
- Las métricas deben validarse teóricamente para dilucidar si miden realmente los atributos que se pretendan medir, lo que permitirá entre otras cosas, conocer la escala de medida que limitará las operaciones posibles con las mediciones.
- Las métricas se deben validar empíricamente, para conocer su grado de utilidad en la obtención de medidas de atributos de calidad externos que no son directamente medibles, o para gestionarlos.
- La obtención de las mediciones debe ser fácil, y a ser posible, automatizada con herramientas adecuadas.

Por la importancia del hecho de medir no se podrá utilizar cualquier sistema de medida, sino que deberá estar sujeto a una teoría: La Teoría de la Medición. El objetivo de una teoría de la medida, es el de lograr que la descripción que las métricas nos aportan del sistema, sea objetiva, exacta, reproducible, segura y significativa.[DMT]

Resultados y Objetivos

Objetivos:

Definir métricas de calidad representativas de los productos de software pertenecientes al dominio del mercado farmacéutico.

Incentivar la discusión académica.

Analizar los modelos y estándares de calidad del software, recorriendo las actualizaciones y nuevos estándares.

Resultados esperados:

Al concluir esta investigación, se pretende brindar métricas que puedan ser consideradas para evaluar productos del dominio mencionado. A fin de que las empresas dispongan de métricas específicas de dicho dominio y se contribuya para mejorar sus productos mediante la retroalimentación de los resultados obtenidos. Tratando de apuntar a la problemática que tienen hoy en día las PYMES que producen software para este mercado y que no figuran como principales proveedoras dentro del mercado farmacéutico.

También se buscará hacer públicas este tipo de métricas, para que sirvan como indicadores de calidad para los grupos empresarios que requieran este tipo de software y puedan invertir con más confianza en nuestro país.

Formación de Recursos Humanos

Entre otras acciones de formación, tesinas de grado y otras tesis de posgrado, se encuentra la de la co-autora Lic. Alejandra Velia Lopez, quien está realizando su tesis de Maestría en Ingeniería de Software de la Universidad Nacional de San Luis, cuyo co-director el

Mg. Ing. Alberto Sánchez y director el Dr. Ing. Germán Montejano.

Referencias

- [FN99] Norman E. Fenton and Martin Neil. A critique of software defect prediction models. IEEE Transactions on Software Engineering, 25(5):675-689, 1999.
- [IEEE05] Computer Society IEEE. Guide to the Software Engineering Body of Knowledge: 2004 Edition - SWEBOK. 2005.
- [Pre05] Roger S. Pressman. Ingeniería del software : un enfoque practico. McGraw-Hill, 6a edition, 2005.
- [PGC07] Piattini, M., García F., Caballero, I. "Calidad de los Sistemas Informáticos". Editorial Alfaomega. 2007.
- Calidad de Sistemas de Información, 2a Edición: Mario Piattini Velthuis, Félix O. García Rubio, Ignacio García de Guzmán y Francisco J. Pino, Alfaomega Grupo Editor, México, Octubre 2011
- Calidad en el desarrollo de software, Pantaleo Guillermo, Alfaomega Grupo Editor Argentino S.A. 2011
- [FP97] Norman Fenton and Shari Lawrence Peeger. Software Metrics: A Rigorous and Practical Approach. International Thomson Computer Press, London, UK, second edition, 1997.

- [Som05] Ian Sommerville. Ingeniera del software. Pearson Educacion, 7a edition, 2005.
- [DMT04] Tesis Doctoral: Metodología para el diseño de métricas en tiempo de ejecución. Universidad de Oviedo. Aquilino Adolfo Juan Fuente, 2004
- Ignacio Pérez y Pedro I Ferrer - Métricas, un enfoque cuantitativo a la gestión de proyectos de software.
- Tesis Doctoral: Deteccion de defectos de diseño mediante metricas de codigo - por Carlos Lopez Nozal. Universidad de Valladolid – Octubre de 2012
- [Piat03] Piattini , García, “Calidad en el desarrollo y mantenimiento del software”, RA-MA Editorial, Madrid, 2003
- [KiP96]Kitchenham, B. y Pfleeger, S. L., “Software Quality: The Elusive Target”, IEEE Software, vol. 20, n° 1, pp. 12-21, 1996
- ISO, “ISO/IEC 9126-1 – Software engineering – Product quality – Part 1: Quality Model”, 2001.
- ISO, “ISO/IEC 9126-2 – Software engineering – Product quality – Part 2: External Metrics”, 2003.
- ISO, “ISO/IEC 9126-3 – Software engineering – Product quality – Part 3: Internal Metrics”, 2003.
- ISO, “ISO/IEC 9126-4 – Software engineering – Product quality – Part 4: Quality in Use Metrics”, 2004
- **Paginas web**
- <http://www.aemes.org/> : Asociación Española para la Gobernanza, la gestión y la medición de las tecnologías de la información.
- <http://japha.org/> : Journal of the American Pharmacists Association
- <https://www.ibm.com/developerworks/ssa/industry/library/ind-openemr/> : Sreevidya Krishna, Programmer and Business Analyst, Freelance 29-07-2011.

Modelado y Análisis Probabilístico de Sistemas Híbridos

Vilallonga, Gabriel^{1,2}; Riesco, Daniel², Sérgio Campos³

¹Departamento de Informática de la Facultad de Ciencias Físico, Matemáticas y Naturales de la Universidad Nacional de San Luis. Ejercito de los Andes 950. San Luis

Tel: +54 (0266) 4520300. Int. 2127

²Departamento Sistemas de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca

Maximio Victoria N° 55 - C.P: 4700 - San Fernando del Valle de Catamarca

Telefono: 03834- 435112 – int 168

³Department of Computer Science, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Pampulha, 30123-970, Belo Horizonte, Brazil.

gvilallo@tecno.unca.edu.ar/unsl.edu.ar, driesco@unsl.edu.ar, scamposg@dcc.ufmg.br

Resumen

Los sistemas híbridos, se han tornado de gran interés en la comunidad científica a partir del desafío que presenta el estudio de sus dinámicas, las continuas y las discretas, y el estudio y comprensión de sus interacciones. Estas puede tomar diversas formas, las más comunes ocurren cuando cambian de estado entre diferentes procesos continuos. Otras formas de interacción incluyen transiciones discretas que dependen de evoluciones continuas, u otras aparecen como resultados de una decisión, o por la ocurrencia de ciertos eventos.

Un sistema híbrido probabilístico considera la distribución de probabilidad de ambas dinámicas, y se enfoca el análisis en lo referido a la alcanzabilidad probabilística. El acercamiento numérico sufre del problema de explosión de estados y son computacionalmente muy exigentes. Un método alternativo de análisis es el realizado por medio de Model Checkers Probabilísticos. En nuestra línea de investigación proponemos el modelado y estudio de estos sistemas, como parte de la verificación y validación de sistemas desde un punto de vista de la ingeniería de software por medio de herramientas de model checking. En una primera etapa se centrará en modelos, específicamente de sistemas biológicos, con el

objetivo de mejorar el poder predictivo de modelos formales existentes.

Palabras claves: Sistemas Híbridos (SH), Sistemas Híbridos Probabilísticos (SHP), Model Checking (MC), Model Checking Probabilístico (MCP), Sistemas Biológicos (SB), Ingeniería de Software (IS).

Contexto

El proyecto de investigación “*Fortalecimiento de la Calidad y la Productividad en Ingeniería de Software*”, posee una línea de investigación dedicada a la verificación y validación de sistemas. Dentro de ella es que se prevé trabajar con SHP.

El mencionado proyecto es evaluado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Catamarca y se desarrolla y ejecuta en ámbitos del Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa y en el Departamento de Informática de la Facultad de Ciencias Físico, Matemáticas y Naturales de la Universidad Nacional de San Luis donde los integrantes se desempeñan como docentes de la carrera de Ingeniería en Informática (UNCa), e Ingeniería en Informática y Licenciatura en Ciencias de la Computación (UNSL) respectivamente. También se cuenta con el convenio establecidos entre la UNSL y la Universidad

Federal de Minas Gerais (UFMG), en el marco del programa Centros Asociados para el Fortalecimiento de Posgrados Brasil/Argentina (CAFP-BA) tanto para maestrías como para doctorados. Por medio de esta se ha logrado generar un nexo con investigadores responsables del LABORATÓRIO DE NEUROCIÊNCIA EXPERIMENTAL E COMPUTACIONAL DR. ARISTIDES AZEVEDO PACHECO LEÃO (LANEC) de la Universidad Federal de Sao João del-Rei, (UFSJ) en lo que respecta a los sistemas biológicos, los cuales se propone estudiar.

Esta línea de investigación, inserta en el proyecto, establece como prioridad la formación científico-técnico de los integrantes con la premisa de proceder a la transferencia de resultados a los medios informáticos como así también al ámbito biológico.

Introducción

Los SH se encuentran en la vida cotidiana, desde una pelota picando, hasta autos sofisticados, trenes, aviones, robots, sistemas biológicos, entre otros. Estos presentan un gran desafío en el área computacional no solo en el estudio de estos, sino que también en las herramientas que se poseen para su modelado y análisis. La coexistencia de la dinámica continua y discreta plantean una mejora a los modelos matemáticos hasta ahora vigentes, como así también de las herramientas de análisis disponibles en la actualidad. Aunque el acercamiento numérico es altamente preciso, este sufre del problema de la explosión de estados y de ser computacionalmente muy exigente.

Las aplicaciones modernas de SH han ido creciendo en complejidad. Esto se da debido a alta interacción, las dinámicas complejas, el comportamiento cambiante, falta de certeza en mediciones y la tolerancia a las fallas. Muchas veces la manera de reducir la complejidad en el estudio de estos tipos de sistemas ha sido empleando modelos probabilísticos. La aleatoriedad se ha convertido en un método estándar en el modelado y análisis de modelos complejos. El hecho es que los SHP [1] en la actualidad son un tópico de gran interés. El

estudio de este tipo de sistemas permite un acercamiento a familias de modelos que resultan de varios tipos de aleatoriedad de sistemas híbridos. Existen SHP con distribuciones de probabilidad asociadas solamente a transiciones discretas. Otros modelos probabilísticos consideran el ruido que perturba la evolución continua. En la forma más general, un SHP considera la distribución de probabilidad para ambos, transiciones discretas y continuas, y, además, estas distribuciones pueden depender de otras.

Los modelos probabilísticos y las lógicas temporales como las lógicas estocásticas continuas y las lógicas arborescentes computacionales probabilísticas son a menudo usados para el modelado y análisis de sistemas con el fin de conocer acerca de su rendimiento y confiabilidad. Existen dos acercamientos al análisis de comportamientos probabilísticos, el numérico y el estadístico.

En el acercamiento numérico el modelo formal del sistema es chequeado (model checked) observando su corrección con respecto a la especificación usando métodos numéricos y simbólicos. Los chequeadores de modelos para las diferentes clases de procesos probabilísticos y especificaciones de lógicas han sido desarrollados destacándose los model checkers probabilísticos como UPPAAL [2-4] y PRISM[5]. En estos tipos de sistemas es imperioso el uso de métodos formales que permitan su estudio para garantizar el correcto desempeño.

Como se menciona en párrafos previos nuestro trabajo, en esta primera etapa, se va a centrar en un caso particular de SH, como lo son los procesos biológicos que exhiben cambios de comportamiento. En etapas posteriores, y en forma gradual, se irán incorporando otros sistemas SHP.

En los SB, la naturaleza estocástica de los procesos celulares ha motivado el uso de SHP para el modelado de estos tipos de fenómenos, ya que combinan las generalidades de SH con procesos probabilísticos [6].

Los SHP han mostrado proveer un marco de trabajo adecuado para el modelado de la

evolución temporal de poblaciones de especies químicas involucradas en conjuntos de reacciones químicas. Las reacciones químicas que se producen dentro de la células son frecuentemente modeladas usando formulaciones probabilísticas, las cuales toman en cuenta la aleatoriedad inherente de movimiento molecular térmico [7]. En estos sistemas las reacciones son tratadas como eventos probabilísticos que cambian la población de las especies individuales basadas en la estequiometría de las reacciones.

Desde el punto de vista de la IS, la complejidad de estos tipos de sistemas desafían la capacidad de las herramientas actuales, donde, de las experiencias realizadas por medio de experimentos sobre las herramientas, pueden surgir sugerencias de mejoras y uso de las experiencias de usuarios resultantes. Este último ítem se basa en que muchas veces los sistemas estudiados pertenecen a áreas muy diferentes a la informática, como por ejemplo la biológica, y el lenguaje de comunicación debe ser revisado para una mejor comprensión entre los investigadores de las distintas disciplinas.

Los resultados en esta línea de investigación pueden mejorar el proceso de formación académica, de docentes y alumnos en el desarrollo de aptitudes y de valores, contribuir en la toma de decisiones tácticas y estratégicas de la Facultad, proporcionando un generador de conocimiento y elevando de esta manera la calidad de la educación que favorezca la formación de ingenieros en la Universidad.

Línea de Investigación, Desarrollo, e Innovación

Inserto en el marco del proyecto se encuentra la línea de investigación referida a la validación y verificación de sistemas. En este caso puntual se hace fuerte hincapié en el área de SHP, donde la mayoría de los casos extra informáticos, han sido llevados a cabo por investigadores de la misma área, sin intervención de investigadores de nuestra disciplina. Esto es sumamente notorio en los

primeros contactos interdisciplinarios.

Esta línea de investigación permite abrir una perspectiva multidisciplinar que tiene como efecto sinérgico el trabajo con sistemas reales, de otras áreas, que permiten la revisión de las herramientas informáticas utilizadas en el área de validación y verificación de sistemas complejos.

Estas actividades han exigido realización de cursos acerca de la temática, como así también el estudio de material bibliográfico y de publicaciones científicas.

Es de hacer notar que estos grupos de trabajos están coordinados por docentes investigadores de las distintas universidades.

Resultados y Objetivos

El objetivo principal de esta línea de investigación es el de contribuir de manera efectiva y mensurable en la concreción de aportes concretos al área de verificación y validación insertas en IS aplicada, lo que implica revisión, o nuevas propuestas, de técnicas, metodologías, y el desarrollo de herramientas que asistan al desarrollo de software específico para el trabajo con SHP de una disciplina determinada, en nuestro caso, en esta primer etapa, en la biología.

El efecto deseado, también, es incidir significativamente en las actividades académicas de grado y posgrado, la formación de recursos humanos, y la transferencia al medio.

Estos objetivos están siendo alcanzado gracias al trabajo conjunto entre los equipo de la UNSL-UNCa, UFMG y UFSJ. El trabajo interdisciplinar está generando aportes significativos gracias a la interacción con científicos del LANEC, referido al área biológica, específicamente en el trabajo con Bombas de Sodio-Potasio afectadas con Palitoxina[8,9].

Formación de Recursos Humanos

Esta línea de investigación está en una etapa intermedia, donde integrantes del proyecto desarrollan sus tesis de doctorado y maestría en el marco de la carrera de doctorado en Ingeniería de Software, como así también de

la maestría en Ingeniería del Software en temas específicamente relacionados a la línea de verificación y validación. Además los participantes pertenecen a distintas cátedras de las carreras de ingeniería y licenciatura de las universidades que participan.

Se ha procedido a la incorporación de alumnos de los últimos años con la finalidad de incluirlos en actividades de investigación y desarrollo en las áreas específicas del proyecto

El proyecto prevé un programa de capacitación y formación de recursos humanos, que contempla las siguientes actividades:

- Dirección de tesinas de grado de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa integrados al proyecto, como así también en la UNSL.
- Participación de los integrantes del proyecto en cursos de actualización y posgrado en el área de estudio.

Para garantizar la capacitación y actualización del equipo de investigación, así como la difusión de los avances y resultados logrados, se propuso la participación en eventos nacionales e internacionales de la especialidad, como congresos, simposios, seminarios y cursos.

Bibliografía

[1] G. Pola, M. L. Bujorianu, J. Lygeros, and M. D. D. Benedetto. Stochastic hybrid models: an overview. Proc. of the IFAC Conference on Analysis and Design of Hybrid Systems, pages 45-50, 2003.

[2] David, A., Larsen, K.G., Mikucionis, M., Poulsen, D.B.. Statistical Model Checking for Stochastic Hybrid Systems. In Proc. HSB 2012. Computational Engineering, Finance, and Science (cs.CE); Software Engineering (cs.SE). arXiv:1208.3856. 2012.

[3] Bulychev, P., David, A., Larsen, K.G., Mikucionis, M., Legay, A., Wang, Z.. UPPAAL-SMC: Statistical Model Checking for Priced Timed Automata. Logic in Computer Science, Formal Languages and Automata Theory. In Proceedings QAPL 2012. DOI:10.4204=EPTCS.85.1

[4] Uppaal in a nutSHPll. Kim G. Larsen, Paul Pettersson, Wang Yi. International Journal on Software Tools for Technology Transfer. December 1997, Volume 1, Issue 1-2, pp 134-152.

[5] Kwiatkowska, M., Norman, G., Parker, D.: PRISM 4.0: Verification of probabilistic real-time systems. In: Proc. CAV. Springer (2011).

[6] Singh, Abhyudai, and Joo P. Hespanha. "Stochastic hybrid systems for studying biochemical processes." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 368.1930 (2010): 4995-5011.

[7] Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. of Comp. Physics, 22, 403-434, 1976.

[8] Rodrigues, A. M., Almeida, A. C. G., Infantosi, A. F., Teixeira, H. Z., & Duarte, M. A. (2008). Model and simulation of Na⁺/K⁺-pump phosphorylation in the presence of palytoxin. Computational biology and chemistry, 32(1), 5-16.

[9] Artigas, P., & Gadsby, D. C. (2003). Na⁺/K⁺-pump ligands modulate gating of palytoxin-induced ion channels. Proceedings of the National Academy of Sciences, 100(2), 501-505.

Cálculo de Métricas para Medir el Grado de Entendimiento de una Descripción WSDL

Hernán Bernardis⁽¹⁾, Edgardo Bernardis⁽¹⁾, Mario M. Berón⁽¹⁾, Daniel E. Riesco⁽¹⁾, Pedro Rangel Henriques⁽²⁾, Maria Joao V. Pereira⁽³⁾

⁽¹⁾Departamento de Informática / Facultad Ciencias Físico Matemáticas y Naturales/ Universidad Nacional de San Luis
Ejército de los Andes 950 – San Luis – Argentina
{hbernardis, ebernardis, mberon, driesco}@unsl.edu.ar

⁽²⁾ Departamento de Informática/Universidade do Minho
Braga – Portugal
pedrorangelhenriques@gmail.com

⁽³⁾ Departamento de Informática e Comunicações/ Instituto Politécnico de Bragança
Bragança - Portugal
mjoao@ipb.pt

Resumen

En la actualidad, existe un gran incremento en el desarrollo de Web Services. Cada vez más organizaciones crean servicios web y los publican en la red con el objetivo de que otras organizaciones o personas los utilicen para construir diferentes sistemas de software. Esta tendencia ha generado que la Ingeniería de Software se dedique en los últimos años a especificar su uso definiendo estándares y herramientas que faciliten la tarea de definición, construcción, implementación y uso de los mismos. Como resultado, actualmente existen estándares aceptados a nivel mundial que especifican el uso de los mismos. Aprovechándose de estos estándares altamente abstractos y formales en su definición, ha sido posible definir herramientas que permiten construir servicios Web de manera casi automática. Sin embargo, esta alta estandarización ha provocado que la comprensión de los Web Services (WS)

sea cada vez más difícil y tediosa. Comprensión que se vuelve fundamental a la hora de determinar de una gran cantidad de Web Services, cuál usar, para realizar tareas de mantenimiento sobre un Web Services o sobre un sistema que invoca múltiples Web Services.

En este artículo se describe una línea de investigación centrada en facilitar la comprensión de Servicios Web mediante el análisis de sus especificaciones WSDL. Análisis que utiliza diferentes estrategias para calcular métricas que permiten determinar cualitativa y cuantitativamente la dificultad de comprensión que presenta un WS.

Palabras clave: Web Services, métricas, comprensión, LSP.

Contexto

La línea de investigación descrita en este artículo se desarrolla en el Laboratorio de Calidad e Ingeniería de

Software (LaCIS) de la Universidad Nacional de San Luis; y se encuentra enmarcada dentro del proyecto: *“Ingeniería del Software: Conceptos, Métodos, Técnicas y Herramientas en un Contexto de Ingeniería de Software en Evolución”*, perteneciente a la universidad antes mencionada. Dicho proyecto, es reconocido por el programa de incentivos, y es la continuación de diferentes proyectos de investigación de gran éxito a nivel nacional e internacional.

Introducción

Actualmente, con el auge de internet se están popularizando los Web Services como artefactos de software a partir de los cuales se pueden construir sistemas más complejos. Según la W3C, un Web Service es: *“Una aplicación de software identificada por una URI, cuya interface y enlaces son capaces de ser definidos, descritos y descubiertos como artefactos XML. Un web service soporta interacción directa con otros agentes de software usando mensajes basados en XML intercambiados a través de protocolos basados en internet”*. Actualmente, muchas organizaciones desarrollan Web Services y los publican en internet con el objetivo de que otras personas u organizaciones (o incluso ellas mismas) los usen para construir sus propios sistemas de software. Esto ha modificado la forma mediante la cual se desarrollan actualmente los sistemas debido a que se amplía el rango de opciones en el desarrollo. Cualquier persona u organización en el mundo puede crear un Web Service y publicarlo en internet. De manera equivalente, este WS puede ser utilizado por cualquier otra persona u organización en el mundo en un sistema propio sin necesidad de saber de qué manera está implementado, ni donde se

encuentra alojado físicamente. Solamente se necesita su especificación para que el agente de software propio la interprete y se comunique con el WS. Luego, se utiliza la información retornada por el mismo en el sistema propio de acuerdo a lo que se necesite hacer.

La idea de poder usar un Web Service desarrollado por cualquier otra persona u organización en un sistema propio ha sido posible debido a la creación de estándares y lenguajes formales de definición de los mismos. Sin embargo, esta alta abstracción en la construcción y especificación de los mismos dificulta en gran medida su comprensión a la hora de realizar tareas de mantenimiento (adaptación, arreglo de errores, migración, etc.) sobre ellos mismos o sobre los sistemas de software que los utilizan.

Todo Web Service posee una especificación que provee la información necesaria para invocarlo. Uno de los estándares de definición más conocido es WSDL (Web Service Definiton Language) [1]. Las especificaciones WSDL son un dialecto XML, con reglas bien definidas para especificar cada componente del WS. Cuántos parámetros recibe y de qué tipo son, qué datos retorna y de qué tipo, qué protocolo de internet usa para su comunicación, qué operaciones posee, son entre otras tantas, características del WS que se encuentran especificadas en su WSDL asociado. Suena lógico entonces comprender un WS mediante el análisis de su correspondiente especificación WSDL.

Aprovechando la alta estandarización presente en el lenguaje WSDL, se pueden definir metodologías de comprensión de los mismos mediante la extracción y análisis de la información presentes en dichas especificaciones [2, 3, 4].

En este trabajo se extrae información aplicando técnicas de compilación, algoritmos de análisis de lenguaje natural y técnicas de cálculo de indicadores sobre su especificación WSDL. Toda esta información se utiliza para definir, calcular y obtener métricas de los WS's que permitan determinar la dificultad de comprensión que poseen [5, 6]. Además, también se utiliza LSP (Logic Scoring of Preference) para definir estructuras de agregación que le asignen pesos a los valores de cada métrica según sea la necesidad del ingeniero de software y, en base a estos pesos, se obtiene un grado de entendimiento global de la especificación WSDL.

La organización de este artículo se expone a continuación. La sección 2 describe la línea de investigación abordada. La sección 3 presenta los resultados obtenidos hasta el momento, junto con aquellos esperados a corto plazo. Finalmente, la sección 4 describe las tareas realizadas por los recursos humanos en formación.

Líneas de Investigación, Desarrollo e Innovación

El análisis de las especificaciones WSDL posee múltiples etapas con su función particular dentro del proceso global de comprensión. En las subsecciones siguientes se describen brevemente dichas etapas.

Extracción de Información

Debido a que las especificaciones WSDL son un dialecto XML, se pueden usar técnicas de compilación sobre las mismas basadas en los parser DOM (Domain Object Model) [7]. Un parser DOM es un parser específico para el lenguaje XML,

el cual construye una representación interna, basada en árboles, del código fuente XML que recibe. Luego, a partir de funciones especialmente diseñadas para recorrer la representación interna construida (funciones transversales) se puede extraer la información deseada. Usando estas funciones transversales, se logran extraer los identificadores de cada componente (*name, type element, etc.*), la documentación y los comentarios presentes en el WSDL.

Cálculo de Métricas

A partir de la información extraída por el parser DOM descrito en la subsección anterior, se procede al cálculo de múltiples métricas. Las métricas definidas en este trabajo de investigación se pueden dividir en dos grupos:

- **Métricas de tamaño:** permiten medir cantidades de componentes de cada tipo, como por ejemplo cantidad de tipos complejos, de parámetros, de operaciones, de mensajes, entre otras. Esto permite tener una idea del tamaño de cada sección particular del WSDL y determinar que tan complejo es, a primera vista, su comprensión. Estas métricas son muy útiles a la hora de contextualizar al ingeniero de software respecto de la dificultad que puede implicar comprender ese WS. También sirven de gran utilidad a la hora de comprender múltiples WS's debido a que fijan valores de comparación. Así, un WS cuya especificación posea 15 tipos complejos posiblemente demande más esfuerzos de comprensión que uno con 2.
- **Métricas de calidad:** permiten medir la calidad semántica de la especificación WSDL. Esto es, que tanta información semántica brinda la especificación WSDL respecto del WS que representa y que tan entendible y

comprensible en sí es dicha especificación. La presencia de identificadores definidos solamente con iniciales, de errores ortográficos y/o gramaticales, de abreviaciones generan complicaciones en la comprensión de las especificaciones WSDL.

Para el cálculo de estas métricas se utilizan algoritmos de procesamiento de lenguaje natural junto con diccionarios de palabras que permiten determinar el grado semántico que poseen los identificadores, comentarios y la documentación presente en la especificación WSDL. Cabe destacar que sobre los identificadores se aplican algoritmos de división y expansión de palabras de manera tal de poder encontrar palabras abreviadas dentro de sus nombres.

LSP

LSP es un método de evaluación multicriterio basado en un árbol de criterios, una estructura de agregación y un conjunto de funciones de criterios elementales definidos [8]. Combinando sistemáticamente estos elementos, este método permite obtener un nivel de satisfacción que indica, en el caso de esta investigación, el grado de entendimiento que posee la especificación WSDL de un WS. Pero, para lograr obtener este grado de entendimiento, se deben definir las subestructuras de evaluación de cada componente dentro de la especificación WSDL. En la Figura 1 se muestra, a modo ilustrativo, la estructura de agregación específica para medir el grado de entendimiento de los mensajes presentes en un WSDL. Dicha estructura evalúa, para cada mensaje, 3 criterios:

calidad de la documentación, calidad del nombre y entendimiento de cada componente part. A cada uno de estos criterios se le asigna un valor numérico obtenido por las métricas definidas en esta investigación. Luego, los criterios son evaluados usando diferentes operadores propios de LSP con diferentes pesos para cada evaluación (pesos que se muestran en los arcos de la estructura de agregación). El lector interesado en profundizar más sobre estos conceptos puede leer [9, 10].

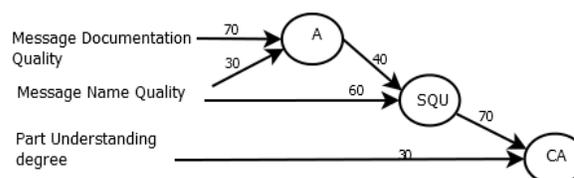


Figura 1. Estructura de agregación para medir el grado de entendimiento de los mensajes.

Resultados y Objetivos

Algunos de los resultados más destacados obtenidos por esta investigación son:

- Se establecieron criterios y métodos para la extracción de información desde las especificaciones WSDL de los WS.
- Se definieron y calcularon diferentes métricas que permiten medir la complejidad de los WS. Dichas métricas miden desde cantidad de componentes de cada tipo presentes en el WSDL como calidad semántica de la especificación mediante el análisis semántico de identificadores, documentación y comentarios existentes en la especificación WSDL.
- Se utilizó LSP para construir una estructura de agregación que permita relacionar los valores de todas las

métricas mediante operadores para obtener un grado de entendimiento global del WS.

- Se construyó WSDLUD, una herramienta que realiza la extracción de información, cálculo de métricas, el cálculo del grado de entendimiento del WS usando LSP y la visualización de la información de múltiples WS de manera automática.

Entre los objetivos planteados a corto y largo plazo se pueden mencionar:

- Detectar y analizar automáticamente relaciones entre WSDLs mediante el estudio de sus espacios de nombres.
- Realizar un análisis más exhaustivo sobre los tipos complejos usados en cada WS.

Formación de Recursos Humanos

Las tareas realizadas en el contexto de la presente línea de investigación están siendo desarrolladas como parte de trabajos para optar al grado de Magister en Ingeniería de Software. En el futuro se piensa generar diferentes tesis de maestría y doctorado a partir de los resultados obtenidos de los trabajos de licenciatura en curso.

Referencias

[1] WSDL Specification for W3C <https://www.w3.org/TR/wsdl>.

[2] N. Gold and K. Bennett. "Program comprehension for web services". In Program Comprehension, 2004. Proc. 12th IEEE International Workshop on. June 2004.

[3] L. O'Brien Lero and D. Smith. "Working session: program comprehension strategies for web service and service oriented archi-

tures". Proc. of 12th IEEE International Workshop on Program Comprehension. 2004.

[4] H. El Bouhissi, M. Malki, and D. Bouchiha. "A reverse engineering approach for the web service modeling ontology specifications". In Sensor Technologies and Applications 2008. SENSORCOMM '08. Second International Conference on, pages 819–823, Aug 2008.

[5] C. Mao. "Towards a data complexity metric set for web service composition". In Computer and Information Technology (CIT), 2011 IEEE 11th International Conference on, pages 127–131, Aug 2011.

[6] Fangfang Liu, Yuliang Shi, Jie Yu, Tianhong Wang, Jingzhe Wu. "Measuring Similarity of Web Services Based on WSDL". IEEE International Conference on Web Services ICWS. 2010.

[7] Parser DOM specification for W3C. <https://www.w3.org/DOM>.

[8] Olsina, L., Rossi, G.: Measuring Web Application Quality with WebQEM. IEEE MultiMedia, 2002 09(4), 20–29 (2002).

[9] Mario M. Berón, Hernán Bernardis, Enrique A. Miranda, Daniel E. Riesco, Maria João Pereira, Pedro Rangel Henriques. "WSDLUD: a Metric to Measure the Understanding Degree of WSDL Descriptions". Proceedings of the 2015 Symposium on Languages, Applications and Technologies, SLATE'15. Madrid, España 2015.

[10] Bernardis, Hernán; Berón Mario; Bernardis, Edgardo; Riesco, Daniel; Henriques, Pedro. "Extracción de información y cálculo de métricas en WSDL 1.1 y 2.0". II Congreso Nacional de Ingeniería Informática / Sistemas de información (CoNaIISI). Argentina. 2014.

Procesamiento de Lenguaje Natural para Estudiar Completitud de Requisitos

Claudia S. Litvak^{1,3}, Graciela D. S. Hadad^{2,3}, Jorge H. Doorn^{1,2}

¹Departamento de Ingeniería e Investigaciones Tecnológicas, UNLaM
Florencio Varela 1903. San Justo. Buenos Aires. 4480-8900

²Escuela de Informática, UNO

³Facultad de Ingeniería y Tecnología Informática, UB
e-mail: claudialitvak@gmail.com, gracielahadad@gmail.com, jdoorn@exa.unicen.edu.ar

Resumen

La Ingeniería de Software incluye el estudio inicial de los requisitos del sistema de software a desarrollar. A ello se dedican los Ingenieros de Requisitos, generando modelos que logren representar las necesidades de los clientes y usuarios y las soluciones a las mismas. El desarrollo de modelos de requisitos en lenguaje natural facilita la interacción entre todos los involucrados, aunque contribuye a generar ambigüedades.

Por otro lado se ha comprobado la existencia de omisiones en una cantidad muy superior a la imaginada intuitivamente. Por su propia naturaleza, las omisiones son notoriamente elusivas. El desarrollo de un mismo modelo acerca del mismo problema por personas diferentes es la forma más eficaz para vislumbrar la magnitud de las omisiones en el dominio de la Ingeniería de Requisitos y posiblemente en muchas otras áreas. Para realizar estudios sobre la completitud de los modelos desarrollados, a fin de mejorar la calidad de los mismos, se ha propuesto la generación de una herramienta de software que permita comparar modelos escritos en lenguaje natural. Se propone que esta herramienta

utilice recursos del Procesamiento del Lenguaje Natural.

Palabras clave: Ingeniería de Requisitos, Completitud de Modelos, Procesamiento de Lenguaje Natural.

Contexto

Esta propuesta está incluida en los proyectos “Control de Calidad de requisitos escritos en Lenguaje Natural” de la Universidad Nacional de la Matanza (UNLaM) y “Tratamiento de los factores situacionales y la completitud en la ingeniería de requisitos” de la Universidad Nacional del Oeste (UNO).

Además esta propuesta es parte del plan de trabajo de becarios alumnos de UNLaM.

Introducción

La Ingeniería de Requisitos tiene como objetivo realizar la definición de los requisitos de un sistema de software, lo más completa posible. Ello implica la mejora de la calidad de los requisitos, dado que la completitud se considera parte de la calidad de los mismos [IEEE 29148]. Diferentes autores han planteado que la incompletitud es uno de los

principales problemas que afronta la Ingeniería de Requisitos [Kotonya 98] [Loucopoulos 95] [Firesmith 05].

Los requisitos se plasman mediante diferentes modelos, en particular, el modelo donde se encuadra este proyecto es el modelo presentado por Leite et al. [Leite95] [Leite97] [Leite00] [Leite01] [Leite04] donde se construyen los requisitos a partir de los modelos Léxico Extendido del Lenguaje (LEL) y Escenarios. Dichos modelos se encuentran escritos en lenguaje natural, facilitando la comprensión e interacción con todos los involucrados. En el caso del modelo LEL, este es un glosario de los términos del dominio de la aplicación, mientras que los escenarios describen situaciones de dicho dominio. En el modelo LEL, para cada término de dicho glosario se definen su nombre, conformado por una palabra o frase, con todos los sinónimos utilizados. El significado del término se lo describe a través de su noción y su impacto. La noción se refiere a la denotación del término, mientras que el impacto describe su connotación en dicho dominio.

Diferentes estudios previos han atacado el tema de la completitud en modelos de requisitos [Doorn 03] [Ridao 06] [Hadad 12] [Hadad 13] [Litvak 12] [Litvak 13a] [Litvak 13b] [Litvak 14] [Litvak 15], pero los progresos obtenidos han sido limitados. En una primera etapa la comparación de los modelos se realizó utilizando una granularidad gruesa, ya que se compararon los nombres de los símbolos. Rápidamente esto permitió observar que los resultados obtenidos habían detectado omisiones significativas, pero simultáneamente se habían creado omisiones aparentes. Esta realidad forzó a reducir la granularidad de las comparaciones a sentencias atómicas. Naturalmente este cambio a su vez desencadenó una escalada exponencial en

las dificultades lingüísticas de las comparaciones. Abordar estas dificultades requiere atender específicamente varios inconvenientes:

i) La presencia de sujetos tácitos: los sujetos tácitos deben ser completados a sujetos presentes, dependiendo del símbolo y del contexto (ver Tabla 1).

Tabla 1. Modificación de oraciones con sujeto tácito

Definición del LEL	Ejemplo
Símbolo	Solicitud de adhesión / Solicitud de ahorro
Noción	Es un formulario que completa el solicitante para ser adherente de un grupo de plan de ahorro.
Oración obtenida sin sujeto tácito	Solicitud de adhesión es un formulario que completa el solicitante para ser adherente de un grupo de plan de ahorro.

Tabla 2. Desdoblamiento de oraciones compuestas

Definición del LEL	Ejemplo
Símbolo	Adherente /No adjudicatario
Noción	Es un solicitante al que la administradora le acepta la solicitud de adhesión.
Oración obtenidas sin sujeto tácito	Adherente /No adjudicatario es un solicitante al que la administradora le acepta la solicitud de adhesión.
Oraciones finales obtenidas	a) Adherente /No adjudicatario es un solicitante.
	b) Administradora acepta la solicitud de adhesión al Adherente.

ii) La presencia de oraciones compuestas, que deben desdoblarse en varias oraciones simples. Si bien se propone al presentar la definición del LEL que se utilicen en la noción y el impacto oraciones simples, en la práctica muchas de ellas suelen ser oraciones compuestas (ver Tabla 2).

iii) La existencia de oraciones mal ubicadas, colocadas en un término al cual no pertenecen.

La propuesta es entonces la generación de una herramienta automática o semiautomática de comparación entre modelos, pensada inicialmente para el modelo LEL.

Líneas de Investigación, Desarrollo e Innovación

Al intentar generar una herramienta semiautomática de comparación de oraciones, resulta imperioso recurrir a estrategias de Procesamiento del Lenguaje Natural [Alonso 05].

Dichas herramientas realizan el reconocimiento del idioma utilizado, la segmentación de oraciones, palabras y secciones, y luego, dependiendo de la herramienta, un análisis morfológico, sintáctico y/o semántico [Alonso 05]. La herramienta a ser construida deberá ser semiautomática, dado que se espera que algunas oraciones no podrán ser apareadas automáticamente y deberán ser comparadas manualmente.

Como trabajo futuro se propone además aplicar esta herramienta para hacer estudios de completitud sobre varios casos de estudio, aplicando técnicas estadísticas a los resultados obtenidos, al comparar los LEL realizados por ingenieros de requisitos independientes. Asimismo, extender la herramienta para la comparación del modelo de Escenarios, e incluso a los requisitos mismos.

Resultados y Objetivos

Hasta la fecha se han estudiado diferentes trabajos y herramientas sobre Procesamiento de Lenguaje Natural [Alonso 05] [Vallez 07] [VilaresFerro 05] [Méndez 99].

Se planifica que la herramienta produzca un documento, con las oraciones apareadas entre ambos LEL y las oraciones de cada uno de los LEL que no pudieron aparearse, facilitando el apareamiento manual de estas últimas.

Se propone, luego de utilizar la herramienta de comparación, realizar estudios estadísticos para evaluar la completitud de los modelos [Wholin 98] [Doorn 03] [Ridao 06] [Hadad 12] [Hadad 13] [Litvak 12] [Litvak 13a] [Litvak 13b] [Litvak 14]. Se espera con ello eliminar las omisiones aparentes y comprender mejor las causas que originan las omisiones con el fin, a su vez, de mejorar las heurísticas de creación y verificación del LEL.

Formación de Recursos Humanos

El proyecto desarrollado en la Universidad Nacional de La Matanza, tiene ya más de 10 años. Está siendo desarrollado por un grupo de investigación del área de Ingeniería de Requisitos, perteneciente al Departamento de Ingeniería e Investigaciones Tecnológicas. Una de las líneas actuales de investigación de este grupo es la completitud.

El proyecto actual de la Universidad Nacional del Oeste, reside en el LITIC, creado en 2014. Cuenta con cinco investigadores, y uno de los temas abordados es la completitud. A mediados de 2015 se han incorporado dos becarios al plantel.

Además, este trabajo forma parte de la Tesis de Doctorado que la Magíster Claudia Silvia Litvak está desarrollando en UNLP, la cual trata sobre la completitud de los modelos de requisitos, denominada “Gestión de la completitud en la Ingeniería de Requisitos”.

Se incorporaron al proyecto de investigación de UNLaM durante 2015

seis alumnos de cuarto año de la carrera Ingeniería Informática de UNLaM. Los mismos son los alumnos Facundo Velazquez Santillán, Ximena Milla, Diego Lo Giudice, Silvina Cáceres y Belén Ortiz. Dadas estas incorporaciones, se han agregado al proyecto un par de objetivos específicos relacionados con el uso de herramientas de Procesamiento de Lenguaje Natural en el análisis y construcción de los documentos del proceso de requisitos. En particular, los alumnos Facundo Velazquez Santillán, Ximena Milla, Diego Lo Giudice fueron seleccionados por la Universidad para recibir Becas UNLaM, que están cumpliendo desde el 1° de septiembre de 2015. La propuesta de las becas involucra que el alumno Facundo Velazquez Santillán, tenga como objetivo la creación de una herramienta informática que facilite la creación de modelos de requisitos de software escritos en lenguaje natural, en particular los modelos LEL y Escenarios. Los alumnos Ximena Milla y Diego Lo Giudice, tienen como objetivo la construcción de una herramienta que permita la comparación de modelos de requisitos escritos en lenguaje natural, en particular para los modelos LEL y Escenarios, a fin de poder facilitar los estudios de completitud. En el estudio de las necesidades de ambas aplicaciones a desarrollar, ha surgido el primordial interés en la investigación de herramientas de Procesamiento de Lenguaje Natural, que puedan aplicarse, en particular aquellas que sean software libre. Las alumnas Silvina Cáceres y Belén Ortiz se han dedicado inicialmente al estudio de las técnicas de Procesamiento de Lenguaje Natural, y están avocadas a la búsqueda y evaluación de herramientas de software libre de Procesamiento de Lenguaje Natural, así como la factibilidad de ser aplicadas en el contexto de investigación

de los becarios, preferentemente de código abierto.

Referencias

[Alonso 05] Alonso i Alemany, L. (2005). Herramientas Libres para Procesamiento del Lenguaje Natural. 5tas Jornadas Regionales de Software Libre, Rosario, Argentina.

<http://www.cs.famaf.unc.edu.ar/~laura/freeNLP>.

[Doorn 03] Doorn, J.H., Ridao, M. (2003). Completitud de Glosarios: Un Estudio Experimental. 6th Workshop on Requirements Engineering Paracicaba, Brasil: Universidade Metodista de Piracicaba, pp. 317-328.

[Firesmith 05] Firesmith D.G. (2005). Are Your Requirements Complete? in Journal of Object Technology, vol. 4, no. 1, January-February 2005, pp. 27-43.

http://www.jot.fm/issues/issue_2005_01/column3

[Hadad 12] Hadad G.D.S., Litvak C.S., Doorn J.H. (2012). Agregando semántica a técnicas de predicción de completitud en modelos de requisitos. WICC 2012 - XIV Workshop de Investigadores en Ciencias de la Computación, anales electrónicos, Universidad Nacional de Misiones, Posadas, ISBN: 978-950-766-082-5, pp. 392-396, <http://hdl.handle.net/10915/18863>.

[Hadad 13] Hadad, G.D.S., Litvak, C.S., Doorn, J.H., Ridao M.N. (2013). Dealing with Completeness in Requirements Engineering. Encyclopedia of Information Science and Technology, Third Edition. Editorial: IGI Global, Mehdi Khosrow-Pour (ed), Information ScienceReference, Hershey, PA, EEUU, 3° edición, cap. 279, pp. 2854-2863, Julio 2014 (10384 p.). ISBN13: 9781466658882.

[IEEE 29148] IEEE Systems and software engineering - Life cycle processes - Requirements engineering. IEEE, Nueva York (2011).

[Kotonya 98] Kotonya G., Sommerville I. (1998). Requirements engineering – processes and techniques, John Wiley & Sons UK.

[Leite 95] Leite, J.C.S.P., Oliveira, A.P.A. (1995). A Client Oriented Requirements Baseline. Second IEEE International Symposium on Requirements Engineering, IEEE Computer Society Press, pp. 108-115.

[Leite 97] Leite, J.C.S.P., Rossi, G., Balaguer, F., Maiorana, V., Kaplan, G., Hadad, G., Oliveros, A.

(1997). Enhancing a Requirements Baseline with Scenarios. *Requirements Engineering Journal*, Springer-Verlag London Ltd., Vol. 2, N° 4, pp. 184-198.

[Leite 00] Leite J.C.S.P., Hadad G.D.S., Doorn J.H., Kaplan G.N. (2000). A Scenario Construction Process. *Requirements Engineering Journal*, ISSN: 0947-3602, Vol.5, N° 1, pp. 38-61, Springer-Verlag London Ltd., Londres, Reino Unido, Julio 2000. <http://link.springer.com/article/10.1007/PL00010342>

[Leite 01] Leite, J.C.S.P. (2001). Gerenciando a Qualidade de Software com Base em Requisitos. *Qualidade de Software: Teoria e Prática*. Prentice-Hall, Rocha A, Maldonado J, Weber K (eds), Cap. 17, pp. 238-246.

[Leite 04] Leite, J.C.S.P., Doorn, J.H., Kaplan, G.N., Hadad, G.D.S., Ridao, M.N. (2004). Defining System Context using Scenarios, en el libro *Perspectives on Software Requirements*, Kluwer Academic Publishers, EEUU, ISBN: 1-4020-7625-8, Cap. 8, pp. 169-199.

[Litvak 12] Litvak C.S., Hadad G.D.S., Doorn J.H. (2012). Un abordaje al problema de completitud en requisitos de software. CACIC 2012 – XVIII Congreso Argentino de Ciencias de la Computación, Universidad Nacional del Sur, Bahía Blanca, ISBN: 978-987-1648-34-4, pp. 827-836, Octubre 2012. <http://hdl.handle.net/10915/23715>.

[Litvak 13a] Litvak, C.S., Hadad, G.D.S., Doorn, J.H. (2013). Correcciones semánticas en métodos de estimación de completitud de modelos en lenguaje natural. 16th Workshop on Requirements Engineering Montevideo, Uruguay: ORT University, pp. 105-117.

[Litvak 13b] Litvak, C.S., Hadad, G.D.S., Doorn, J.H. (2013). Mejoras semánticas para estimar la completitud de modelos en lenguaje natural. CoNaIISI 2013, 1er Congreso Nacional en Ingeniería Informática / Sistemas de Información. Córdoba, Argentina: Universidad Tecnológica Nacional.

[Litvak 14] Litvak C.S., Hadad G.D.S., Doorn J.H. (2014). Heurísticas para el modelado de requisitos escritos en lenguaje natural. CACIC 2014 - XX Congreso Argentino de Ciencias de la Computación, Universidad Nacional de La Matanza, Buenos Aires, ISBN: 978-987-3806-05-6, Octubre 2014, <http://sedici.unlp.edu.ar/handle/10915/42337>.

[Litvak 15] Litvak C.S., Hadad G.D.S, Doorn J.H. (2015). Influencia de las nominalizaciones sobre la

completitud de modelos de requisitos. WICC 2015 - XVII Workshop de Investigadores en Ciencias de la Computación, anales electrónicos, Universidad Nacional de Salta, Salta, ISBN: 978-987-633-134-0,

[Loucopoulos 95] Loucopoulos P., Karakostas V. (1995). *System requirements engineering*. McGraw-Hill, Book company Europe.

[Méndez 99] Méndez, E., Moreira González, J.A. (1999). Lenguaje natural e indización automatizada. *Ciencias de la Información*, 30(3), pp. 11-24.

[Ridao 06] Ridao, M.N., Doorn, J.H. (2006). Estimación de Completitud en Modelos de Requisitos Basados en Lenguaje Natural. 9th Workshop on Requirements Engineering. Rio de Janeiro, Brasil: Pontificia Universidade Católica do Rio de Janeiro, pp. 151-158.

[Vallez 07] Vallez M. y Pedraza-Jimenez R. (2007). El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines [en línea]. "Hipertext.net", núm. 5. <https://www.upf.edu/hipertextnet/numero-5/pln.html>.

[VilaresFerro 05] Vilares Ferro, J. (2005). Aplicaciones del procesamiento del lenguaje natural en la recuperación de información en español. <http://coleweb.dc.fi.udc.es/cole/library/ps/Vil2005a.pdf>

[Wohlin 98] Wohlin, C., Runeson, P. (1998). Defect content estimations from Review Data. 20th Intl Conference on Software Engineering, Japón, pp. 400-409.

Integración de Metodologías Ágiles y Arquitecturas de Software en el desarrollo de Sistemas de Información

Mg. Mirta E. Navarro¹, Mg. Marcelo P. Moreno², Lic. Juan Aranda³, Lic. Lorena Parra⁴, Lic. Jose R. Rueda⁵

Departamento de Informática - F.C.E.F. y N. - U.N.S.J.

Complejo Islas Malvinas. Cereceto y Meglioli. 5400. Rivadavia. San Juan

¹mirtaenavarro@yahoo.com.ar ²mpmoren@gmail.com ³juanaranda@live.com

⁴lorenaparra152@yahoo.com.ar ⁵joserocardoruada@hotmail.com

Resumen

La Ingeniería de Software (IS) se ha convertido en imprescindible en el ámbito organizacional, principalmente debido a que la mayoría de los proyectos que impulsan la competitividad, depende en gran medida de los Sistemas de Información (SI) y de las Tecnologías de la Información (TI).

En los últimos años, se ha impuesto el uso de las Metodologías Ágiles, marcando una tendencia en su adopción al desarrollo de proyectos de software. La causa principal es que en ambientes donde las necesidades de las organizaciones y la tecnología cambian rápidamente, las metodologías tradicionales predictivas han demostrado ser poco eficientes para atender los requerimientos de clientes y usuarios, limitando la competitividad y a la obtención de mayores beneficios en la producción de bienes o en la prestación de servicios, en el menor tiempo posible.

La Arquitectura de Software, en tanto, se refiere a la estructura que compone un sistema de software, la cual comprende elementos de software, las propiedades externamente visibles de aquellos elementos y su interrelación para satisfacer la funcionalidad y requerimientos deseados. Si bien en los últimos años, ha comenzado a cobrar una mayor importancia dentro de los estudios propiciados en la IS, el desarrollo de la

Arquitectura de Software es una práctica poco común para algunos desarrolladores, en especial si se utilizan metodologías de desarrollo ágiles, en las que esas actividades no se consideran relevantes.

En este trabajo se presenta una línea de investigación tendiente a indagar estrategias posibles de integración entre Arquitectura de Software y Metodologías Ágiles, sin dejar de lado las consideraciones de riesgo y costos asociados a la integración pretendida.

Palabras clave: *Software Architecture, Agile methodologies, Information Systems*

Contexto

El presente trabajo se encuadra dentro del área de I/D de la IS y de los SI, y describe los lineamientos generales del proyecto de investigación: “*Integración de Metodologías Ágiles y Arquitecturas de software en el Desarrollo de Sistemas de Información*”, iniciado en Enero de 2016, con una duración de dos años y que tiene como unidades ejecutoras al Departamento e Instituto de Informática de la FCEFyN de la UNSJ.

El grupo de investigación tiene una trayectoria de 14 años en diferentes proyectos vinculados a Metodologías de Desarrollo y Tecnologías, con numerosas publicaciones en diferentes ámbitos, y con la formación de recursos humanos en el área de interés.

Introducción

Las actuales características dinámicas y variables de las organizaciones, están reconsiderando las bases sobre las que se sustenta el desarrollo software. La IS se ha convertido en una disciplina imprescindible debido a que la mayoría de los proyectos que impulsan la competitividad dependen en gran medida de los SI y de las TI. Para el desarrollo de software existen numerosas propuestas metodológicas que inciden en distintas dimensiones del proceso. Hay metodologías tradicionales que se centran especialmente en el control del proceso, estableciendo rigurosamente las actividades involucradas, artefactos que se deben producir y herramientas y notaciones que se usarán. El resultado final es un proceso de desarrollo complejo. Las metodologías tradicionales han demostrado ser efectivas y necesarias en un gran número de proyectos pero, sobre todo donde el entorno del sistema es muy dinámico, donde las necesidades de las organizaciones y la tecnología cambian rápidamente, han presentado problemas [1]. Ante esas dificultades las Metodologías Ágiles emergen como una posible respuesta para llenar ese vacío metodológico [2], y en donde se exige reducir drásticamente los tiempos de desarrollo, pero manteniendo una alta calidad. Las Metodologías Ágiles aparecen como una opción atractiva pues, en contraposición con las metodologías convencionales que actúan basadas en principios de estabilidad y control del contexto, las ágiles no se centran en habilidades de predicción, ni pretenden tener un sistema perfectamente definido como paso previo a su construcción, sino que perciben cada respuesta al cambio como una oportunidad para mejorar el sistema e incrementar la satisfacción del

cliente, considerando la gestión de cambios como un aspecto inherente al propio proceso de desarrollo software y, permitiendo de este modo, una mejor adaptación al entorno, maximizando la inversión y reduciendo los costos, ya sea para variar parte de su funcionalidad, añadir otra nueva, o por ejemplo, adaptar el sistema a un nuevo dominio de aplicación.

Por otra parte, es importante el análisis de la Arquitectura de Software, que es el referente principal de la estructura que compone un sistema: elementos de software, las propiedades externamente visibles de esos elementos y la interrelación entre ellos. Este tema ha comenzado a cobrar importancia dentro de los estudios propiciados en la IS [3]. La Arquitectura de Software ha sido aplicada exitosamente en conjunto con las metodologías predictivas o tradicionales, permitiendo introducir criterios para la obtención productos de calidad [4]. A pesar de ello, el desarrollo de la Arquitectura de Software continúa siendo una práctica poco común, y en general se considera como un elemento no relacionado a las metodologías ágiles. De hecho, el manifiesto ágil no es muy claro en su exposición frente a la arquitectura, dejando aspectos de diseño librados a la etapa de codificación [5], considerando a la Arquitectura como un complemento.

Aquí es entonces donde se debe reflexionar acerca de cuál sería el paradigma más efectivo para establecer vínculos de integración de la Arquitectura al desarrollo con Metodologías Ágiles y de los beneficios resultantes de esa integración. La línea de investigación que se propone, trabajará en esa dirección, sin descuidar aspectos de riesgo y costos asociados a la integración que se pretende. Como resultado final se pretende la posibilidad de determinar un modelo que pueda integrar actividades de

diseño de Arquitectura de Software en Metodologías Ágiles, considerando que tal modelo no debe contraponerse con los valores y principios fijados en el manifiesto ágil expresados en [5].

Metodologías Ágiles

Las Metodologías Ágiles enfatizan en la colaboración con el usuario, requiriendo que se involucre e interactúe en forma directa con el equipo de desarrollo; en el producto software; y en la respuesta al cambio. Por lo tanto, es una metodología iterativa, orientada hacia prestaciones y hacia la entrega, en detrimento de aspectos relacionados a la documentación y el seguimiento de planes [6]. Por ello aparece como alternativa atractiva para adaptarse a contextos donde los requisitos cambian rápidamente. Estos aspectos hacen que las metodologías ágiles promueven prácticas adaptativas en lugar de predictivas.

Estas metodologías constituyen una solución con una elevada simplificación que a pesar de ello no renuncia a las prácticas esenciales para asegurar la calidad del producto. Por lo tanto, para ejecutar un desarrollo exitoso con metodologías ágiles, es importante comprender los requisitos del cliente, intentando extraer al máximo los deseos del usuario para entregar un producto lo más cercano a la realidad. Al ser flexibles y colaborativas, se ajustan fácilmente a cambios de requerimientos, de hecho, el cambio en los requerimientos es una característica esperada y deseada, al igual que las entregas constantes al cliente y la retroalimentación por parte de él. Tanto el producto como el proceso son mejorados frecuentemente y la capacidad de entrega continua y en plazos breves permite al cliente verificar in situ el desarrollo del proyecto y familiarizarse con la funcionalidad del producto progresivamente y comprobando si el

sistema cumple con los requerimientos, mejorando de esta forma su satisfacción. Además, el desarrollo en ciclos de corta duración favorece que los riesgos y dificultades se repartan a lo largo del desarrollo del producto, principalmente al comienzo del mismo y permite analizar los riesgos y dificultades.

Arquitectura de Software

La Arquitectura de Software es el conjunto de decisiones de estructura de diseño de alto nivel que reúne todos los requerimientos técnicos y operacionales y que son difíciles de cambiar durante el proceso de desarrollo. Esa estructura se compone de elementos de software, las relaciones entre ellos, las propiedades de ambos (elementos y relaciones) y de patrones arquitectónicos que guían a esta organización. Estas incumbencias van más allá de los algoritmos y estructuras de datos de los programas. La arquitectura es principalmente importante para satisfacer los requisitos no funcionales, que están relacionados a los atributos de calidad como el rendimiento, seguridad y escalabilidad.

Hay distintas definiciones aceptadas para la Arquitectura de Software todas varían según su enfoque, que en general van de los técnicos, hasta aquellos que se centran en función del costo de hacer cambios a decisiones de diseño. Desde un enfoque técnico, se puede tomar la especificación [7], que define a la Arquitectura de software como:

“los conceptos o propiedades fundamentales de un sistema en su entorno plasmado en sus elementos, relaciones, y en los principios de su diseño y evolución”.

Los enfoques basados en el costo de cambiar decisiones de diseño centralizan el interés en que las decisiones de arquitectura son significativas, ya que alcanzan efectos amplios o generales, y

por lo tanto, la corrección de errores requiere grandes cantidades de esfuerzo. Esto implica que las consideraciones de arquitectura deben hacerse lo antes posible y que los cambios en la arquitectura generan costos que deben evitarse o minimizarse. En ese sentido, M. Fowler [8] define:

"La arquitectura es el conjunto de las decisiones de diseño que deben hacerse en las etapas tempranas de proyecto".

Las definiciones citadas reconocen que las decisiones arquitectónicas se deben hacer en etapas tempranas, puesto que son difíciles de cambiar, típicamente antes de que comience el desarrollo de los requisitos funcionales. En ese sentido, las consideraciones de los costos de cambios es un aspecto especialmente relevante en el uso de metodologías de desarrollo ágil donde una definición que considera el impacto económico de la arquitectura es muy útil.

Los componentes fundamentales en la Arquitectura de Software, son los patrones arquitectónicos, que se definen como ideas que han sido útiles en un contexto práctico y que probablemente serán útiles también para otros contextos similares. Los patrones son un conjunto de principios que proporcionan un marco abstracto para una familia de sistemas. Un patrón arquitectónico mejora la partición y promueve la reutilización del diseño, proporcionando soluciones a problemas recurrentes con frecuencia. Se puede pensar a los patrones como conjuntos de principios que dan forma a una aplicación y por lo tanto ayudan a definir las características básicas y de comportamiento de una aplicación. Es necesario entonces conocer claramente las características, fortalezas y debilidades de cada patrón de arquitectura, con el fin de elegir el que se adapte a las necesidades y objetivos de negocio específicos. Aunque

hay muchos otros, los patrones pueden ser categorizados en los siguientes tipos:

- Cliente/Servidor
- Arquitectura en Capas.
- Arquitectura de Componentes.
- Arquitectura Pipeline de Mensajes.
- Arquitectura de Servicios (SOA).

Es importante destacar que los patrones describen diferentes aspectos de las aplicaciones. Por ejemplo, algunos patrones describen modelos de despliegue, otros describen aspectos de estructura y diseño, y otros describen factores de comunicación. Por lo tanto, una aplicación típica puede utilizar una combinación de más de uno de los patrones mencionados.

Tópico de investigación, Desarrollo e Innovación

Los posibles beneficios de una integración de Arquitectura de Software en las Metodologías Ágiles no es un tema lo suficientemente explorado en ambientes académicos, habiéndose encontrado escasa cantidad de publicaciones relacionadas.

Por otro lado, el interés de indagar en posibles estrategias de integración y posterior evaluación de riesgos/costos, ya había surgido a partir de inquietudes emergentes en proyectos de investigación anteriores que este grupo llevó a cabo [9] [10], siempre siguiendo la línea de las Metodologías de desarrollo, en especial de las ágiles.

Resultados y Objetivos

Los resultados que se esperan obtener al finalizar la investigación son:

- Analizar características y procesos de las Metodologías Ágiles y de la Arquitectura de Software.
- Proponer un modelo de Arquitectura de Software sin contraponerse con las

postulaciones del manifiesto de las metodologías Ágiles.

-Aplicar la integración de Metodología Ágil y la Arquitectura de Software propuesta al desarrollo de diferentes SI de pequeñas organizaciones.

-Creación de un marco de referencia que permita analizar riesgos y estimar costos de la implementación de SI guiada por la integración propuesta.

El objetivo principal es establecer la incorporación de la Arquitectura de Software al proceso de desarrollo dirigido por metodologías ágiles, con el propósito de que esa integración, sin contraposición con los criterios, valores y principios del dominio de aplicación agilista, favorezca el diseño de Sistemas de Información aprovechando las ventajas propias de la Arquitectura de Software, tendiendo a lograr un equilibrio en la integración y reducir la exposición al riesgo.

Formación de Recursos Humanos

El equipo de trabajo está compuesto por cinco docentes-investigadores, y seis alumnos adscriptos. En el periodo 2014-2015 se han asesorado: una tesis de maestría aprobada y finalizada, una tesis de maestría en proceso, dos tesis de grado finalizadas y cinco tesis de grado en proceso, tres de las cuales se están desarrollando en el nuevo tópico de investigación que se presenta en este paper.

Además, se espera realizar direcciones, a los alumnos adscriptos, en diversos trabajos orientados al desarrollo de SI con metodologías iconix-scrum, adoptando el enfoque de integración con Arquitecturas.

Con los resultados de la presente investigación, se harán actividades de divulgación en publicaciones y presentaciones en eventos nacionales e internacionales y también en cursos de postgrado y actualización.

Referencias

- [1] Glass, R.L. Agile Versus traditional: Make Love, not War IT Journal 2011.
- [2] Canós J., Letelier P. “Metodologías Ágiles en el Desarrollo de Software”. JISBD 2003. España.
- [3] Urquiza Yllescas, J.F., et al, “Las Metodologías Ágiles y las Arquitecturas de Software”. Coloquio Nacional de Investigación en Ing. De Sofá. 2010, León, México.
- [4] Evans, E. “Domain-Driven Design” Addison-Wesley 2011 ISBN 0321125215.
- [5] Manifiesto for Agile Software Development Kent Beck, Mike Beedle, et al. www.agilemanifesto.org/ Accedido el 22 de Febrero de 2016.
- [6] Breivold, H.P., Sundmark, D., Wallin, P. and Larsson, S., “What Does Research Say about Agile and Architecture?” , en Proceedings of the Fifth International Conference on Software Engineering Advances (ICSEA), USA, (2010).
- [7] ISO/IEC/IEEE42010. IEEE Std 1471:2000, “Recommended Practice for Architectural Description of Software-intensive Systems”.
- [8] Martin Fowler. “Patterns of Enterprise Application Architecture” Addison-Wesley. 2003 1st Edition. ISBN-13: 007-6092019909.
- [9] “Aplicabilidad de Metodologías y Tecnologías en el desarrollo de Sistemas de Información” Cod: 21/E 979. FCFN. Navarro, Mirta, et al. 2014-2015
- [10] “Convergencia de Tecnologías Informáticas y Metodologías para la Implementación de Sistemas de Información” Cod: 21/E/871 FCFN. Navarro, Mirta, et al. 2012-2013.

AVANCES EN LA CONSTRUCCIÓN DE UN SISTEMA DE RECUPERACIÓN DE INFORMACIÓN PARA INFORMACIÓN CIENTÍFICA EN CIENCIAS DE LA COMPUTACIÓN

H. Kuna¹, M. Rey¹, E. Martini¹, A. Canteros¹, A. Cantero¹, A. Rambo¹, C. Biale¹, E. Zamudio¹

1. Depto. de Informática, Facultad de Ciencias Exactas Quím. y Naturales, Universidad Nacional de Misiones.

{hdkuna}@gmail.com

RESUMEN

La búsqueda de información en internet es una actividad en donde la relevancia de los resultados cobra una importancia cada vez mayor a medida que la cantidad, variedad y disponibilidad de los datos aumentan. En los últimos años han aparecido herramientas de recuperación de información para diferentes ámbitos, constituyendo aplicaciones para el procesamiento de datos de un área en particular, integrando métodos de evaluación de la relevancia adaptados a su contexto, inclusive para el área de Ciencias de la Computación.

La recuperación de información científica se considera un área en la cual el desarrollo de herramientas de este tipo se podría considerar de gran utilidad en la medida que provean de información confiable y de diversas fuentes a un usuario-investigador. La posibilidad de que los resultados a mostrar al usuario integren datos de publicaciones, autores y medios de publicación, enriquecidos a través de metadatos relacionados al impacto generado con sus aportes al tema de búsqueda; además de establecer el orden de presentación a partir de métodos de evaluación con un enfoque multi-dimensional e integral con base en ese mismo impacto, se define como el objetivo principal de la presente línea de investigación.

Palabras clave: *información científica, meta-buscador, producción científica, evaluación de investigación, métricas.*

CONTEXTO

Está línea de investigación articula el “Programa de Investigación en Computación” de la Facultad de Ciencias Exactas Químicas y Naturales (FCEQyN) de la Universidad Nacional de Misiones (UNaM) con el Grupo de Investigación Soft Management of Internet and Learning (SMILE) de la Universidad de Castilla-La Mancha, España.

1 INTRODUCCION

1.1 Sistemas de Recuperación de Información e Información Científica

Los Sistemas de Recuperación de Información (SRI) se reconocen como herramientas de gran utilidad para la gestión de información en internet [1], [2]. Actividades como el almacenamiento, representación, análisis y mantenimiento de grandes volúmenes de datos son llevadas a cabo por herramientas de este tipo en diferentes contextos de aplicación [3]–[5].

En los últimos años, se han generado y publicado numerosos indicadores definidos para la evaluación del impacto generado por publicaciones científicas [6]. Esta situación ha conllevado al desarrollo de herramientas especializadas, con similares funciones a las de los SRI, en la recopilación de información y aplicación de uno o más de tales indicadores para la evaluación de un documento en particular, además de sus autores y los medios seleccionados para su divulgación [7]–[9].

Sin embargo, solo una fracción de los indicadores y las herramientas que los utilizan lo hacen en forma complementaria, es decir que no es habitual poder ejecutar búsquedas de contenido de utilidad para un

investigador científico haciendo uso de la información obtenida a partir del cálculo de indicadores sobre el objetivo de la búsqueda [10]–[12]. Por otra parte, el tipo de evaluación predominante en las soluciones disponibles se basa en el análisis unidimensional de la producción científica, sin considerar la interacción entre indicadores para la evaluación de más de una característica de una publicación en un momento determinado [13], [14].

1.2 Antecedentes

En los últimos años, se han producido avances en el desarrollo de diversos componentes de un SRI de propósito específico para la recuperación de publicaciones científicas del área de Ciencias de la Computación (CC) [15]. Se ha logrado desarrollar un SRI funcional con módulos que permiten su aplicación, entre los que se destacan el módulo de expansión de consultas [16], [17] y el de aplicación de un algoritmo de ranking para el ordenamiento de los resultados a presentar al usuario [18], [19], inclusive se han planteado diferencias en cuanto a los métodos de estimación del impacto de la producción científica [20].

El mencionado algoritmo de ranking ha sido diseñado considerando las principales limitaciones de los modelos de evaluación de la actualidad [8], [21], [22] y, a través de su aplicación, se evalúa a una publicación científica a partir de los valores de diversas métricas, provenientes de fuentes diferentes y que evalúan a cada documento desde tres propiedades o dimensiones: la calidad de la fuente de su publicación, la calidad de sus autores y la calidad del documento en sí mismo [19].

1.3 Inconvenientes detectados en el SRI desarrollado

El SRI desarrollado no se encuentra exento de limitaciones o inconvenientes, entre los que se pueden mencionar:

- Los elementos a recuperar desde las diferentes fuentes a las que accede

el SRI se limitan a publicaciones científicas.

- Las bases de datos sobre las cuales se replican las consultas se corresponden únicamente a buscadores de librerías digitales.
- Los resultados recuperados de cada fuente de datos no son re-analizados una vez que se han ordenado según los criterios del algoritmo de ranking y se han presentado al usuario. De esta manera se descartan sus meta-datos al finalizar el proceso de búsqueda.
- No se mantienen datos correspondientes al feedback del usuario al completar la operación con el SRI, de modo que no se puede brindar una experiencia personalizada en base a sus preferencias de búsqueda.
- No se cuenta con datos históricos correspondientes a búsquedas ejecutadas, lo que dificulta la aplicación y/o generación de nuevos indicadores para la evaluación de resultados en base a un historial de clasificaciones realizadas.

A partir de estos problemas se han planteado cambios en la arquitectura y en el funcionamiento general del SRI, principalmente con el objetivo de recuperar una mayor cantidad de información, incluyendo no solo documentos científicos sino también datos de autores y de fuentes de publicación relevantes para la consulta ingresada por el usuario. De esta manera, se pretende desarrollar un SRI para la recuperación de información científica en un sentido más general para las CC.

2 LÍNEAS DE INVESTIGACION, DESARROLLO E INNOVACIÓN

La búsqueda de información en internet es una actividad habitual para un investigador científico. El uso de buscadores y herramientas especializadas para la recuperación de contenido podría

considerarse de suma utilidad dados los volúmenes de información disponibles en la actualidad. Las CC son consideradas como un área en constante expansión e innovación, por lo que la premisa anterior cobra mayor relevancia.

Es en este contexto que se reconoce como objetivo de interés para la presente línea de investigación, el desarrollo de un SRI que facilite a un usuario investigador la ejecución de búsquedas en forma transparente hacia diferentes fuentes, recuperando no solo datos de documentos, sino también de autores y medios de publicación relacionados. Por otra parte, el mencionado SRI deberá contar con componentes que permitan establecer el orden de los resultados a partir de un análisis multi-dimensional de cada elemento recuperado, integrando toda la información disponible; y previendo el almacenamiento de diversos datos de cada operación con el objetivo de mejorar la experiencia de un usuario en base a resultados calificados como relevantes en búsquedas previas.

3 RESULTADOS Y OBJETIVOS

3.1 Objetivos de la investigación

Dados los inconvenientes detectados, en la presente línea de investigación se propone la refactorización del SRI presentado anteriormente, integrando al mismo la recuperación de información científica de autores y medios de publicación, además de los documentos que se obtienen en la actualidad. Actualizando el método de evaluación de cada tipo de resultado con base en el modelo de evaluación planteado previamente, al contar con mayor cantidad de información.

Además se propone la modificación de la arquitectura planteada para el SRI a fin de incorporar un middleware para la gestión de la recuperación de datos, que se base en las bases de datos a las que accede el SRI, y no en el tipo de resultado a recuperar. De esta manera, se facilitaría la integración de fuentes que no se correspondan únicamente a bibliotecas digitales, pudiendo acceder a

una fuente específica para recuperar un tipo de resultado específico, como podría ser de autores.

El cumplimiento de ambos objetivos permitiría el procesado de datos de documentos, autores y fuentes de publicación tanto para su presentación al usuario, ordenados según el algoritmo de ranking, como así también su post-procesamiento para contar con un mayor volumen de datos para mejorar la calidad de la evaluación de futuras consultas. De igual manera, se podrían persistir los resultados y su evaluación a modo de historial operativo del SRI, pudiendo agregarse datos correspondientes al feedback del usuario.

Con estos cambios, se considera que el SRI constituiría una herramienta de mayor calidad para la presentación de información a un investigador científico.

3.2 Actividades en curso

En el marco de la presente línea de investigación se encuentran en ejecución las siguientes actividades:

- Refactorización del código fuente del SRI para la incorporación en el mismo de resultados del tipo autor y fuente de publicación.
- Desarrollo del middleware para la gestión de la recuperación de datos, definiendo elementos específicos para cada fuente, variando el tipo de resultado a recuperar.
- Desarrollo del componente encargado de la persistencia de los resultados de las búsquedas ejecutadas para su posterior procesamiento.
- Adaptación de los componentes de la interfaz gráfica del SRI para la presentación de los tipos de resultados integrados en esta iteración.
- Diseño e implementación de los perfiles de datos para la persistencia de documentos científicos, autores y fuentes de publicación. Para ello se

encuentra en curso un análisis de las diferentes fuentes de información a fin de extraer los atributos comunes a las mismas y unificando los formatos a utilizar para su implementación.

3.3 Objetivos a corto plazo

A corto plazo los objetivos de la presente investigación son:

- Adaptar los métodos de aplicación del algoritmo de ranking del SRI para la evaluación de los diferentes tipos de resultados. Incorporando en estas operaciones las métricas que pudieran utilizarse a partir de los meta-datos obtenidos de cada resultado a partir de la ejecución de las búsquedas.
- Unificar una taxonomía para las sub-áreas incluidas dentro de las CC a fin de proponer al usuario la selección de una de ellas para la orientación de la búsqueda hacia resultados de mayor relevancia dentro de la misma.
- Diseñar y desarrollar un componente que permita al SRI la gestión de perfiles de usuario, incluyendo un feedback sobre los resultados de las búsquedas que ejecute y parametrizaciones específicas orientadas a sus áreas de mayor interés.

4 FORMACION DE RECURSOS HUMANOS

Este proyecto es parte de las líneas de investigación del “Programa de Investigación en Computación” de la FCEQyN de la UNaM, con diez integrantes (todos ellos alumnos, docentes y egresados de la carrera de Licenciatura en Sistemas de Información de la FCEQyN – UNaM) de los cuales tres están realizando su tesis de grado, cuatro se encuentran realizando una maestría y uno se encuentra realizando un doctorado. La línea y el equipo de

investigación se vinculan con el Grupo de Investigación Soft Management of Internet and Learning (SMILe) de la Universidad de Castilla-La Mancha, España.

5 BIBLIOGRAFIA

- [1] J. A. Olivas, *Búsqueda Eficaz de Información en la Web*. La Plata, Buenos Aires, Argentina: Editorial de la Universidad Nacional de La Plata (EDUNLP), 2011.
- [2] R. Baeza-Yates y B. Ribeiro-Neto, *Modern information retrieval*, vol. 463. ACM press New York., 1999.
- [3] «Semantic Scholar — Allen Institute for Artificial Intelligence». [En línea]. Disponible en: <http://allenai.org/semantic-scholar.html>. [Accedido: 03-mar-2016].
- [4] J. Serrano-Guerrero, F. P. Romero, J. A. Olivas, y J. de la Mata, «BUDI: Architecture for fuzzy search in documental repositories», *Mathware & Soft Computing*, vol. 16, n.º 1, pp. 71–85, 2009.
- [5] J. de la Mata, J. A. Olivas, y J. Serrano-Guerrero, «Overview of an Agent Based Search Engine Architecture», en *Proc. Of the Int. Conf. On Artificial Intelligence IC-AI'04*, Las Vegas, USA, 2004, vol. I, pp. 62-67.
- [6] J. Bollen, H. Van de Sompel, A. Hagberg, y R. Chute, «A Principal Component Analysis of 39 Scientific Impact Measures», *PLoS ONE*, vol. 4, n.º 6, p. e6022, jun. 2009.
- [7] A. N. Guz y J. J. Rushchitsky, «Scopus: A system for the evaluation of scientific journals», *Int Appl Mech*, vol. 45, n.º 4, pp. 351-362, abr. 2009.
- [8] P. Jacso, «As we may search-Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases», *CURRENT SCIENCE-BANGALORE*-, vol. 89, n.º 9, p. 1537, 2005.

- [9] L. I. Meho y K. Yang, «A New Era in Citation and Bibliometric Analyses: Web of Science, Scopus, and Google Scholar», arXiv e-print cs/0612132, dic. 2006.
- [10] M. E. Falagas, E. I. Pitsouni, G. A. Malietzis, y G. Pappas, «Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses», *FASEB J*, vol. 22, n.º 2, pp. 338-342, ene. 2008.
- [11] L. Leydesdorff, «How are new citation-based journal indicators adding to the bibliometric toolbox?», *J. Am. Soc. Inf. Sci.*, vol. 60, n.º 7, pp. 1327-1336, jul. 2009.
- [12] Y. Ding, G. G. Chowdhury, S. Foo, y W. Qian, «Bibliometric information retrieval system (BIRS): A web search interface utilizing bibliometric research results», *J. Am. Soc. Inf. Sci.*, vol. 51, n.º 13, pp. 1190-1204, ene. 2000.
- [13] R. Monastersky, «The number that's devouring science», *Chronicle of Higher Education*, vol. 52, n.º 8, p. 14, 2005.
- [14] J. Ewing, «Measuring journals», *NOTICES-AMERICAN MATHEMATICAL SOCIETY*, vol. 53, n.º 9, p. 1049, 2006.
- [15] H. Kuna, M. Rey, E. Martini, L. Solonezen, y L. Podkowa, «Desarrollo de un Sistema de Recuperación de Información para Publicaciones Científicas del Área de Ciencias de la Computación», *Revista Latinoamericana de Ingeniería de Software*, vol. 2, n.º 2, pp. 107-114, 2013.
- [16] M. Rey, H. D. Kuna, E. Martini, L. Podkowa, J. G. A. Pautsch, y E. Zamudio, «Generación de un método de expansión de consultas basado en ontologías para un sistema de recuperación de información», presentado en XX Congreso Argentino de Ciencias de la Computación (Buenos Aires, 2014), 2014.
- [17] H. D. Kuna, M. Rey, L. Podkowa, E. Martini, y L. Solonezen, «Expansión de consultas basada en ontologías para un sistema de recuperación de información», presentado en XVI Workshop de Investigadores en Ciencias de la Computación, 2014.
- [18] H. Kuna, M. Rey, J. Cortes, E. Martini, y L. Solonezen, «Generating a Ranking Algorithm for Scientific Documents in the Computing Science Area», en *XIX Argentine Congress of Computer Science Selected Papers*, La Plata, Buenos Aires, Argentina: EDULP, 2014, pp. 185-195.
- [19] H. Kuna, E. Martini, y M. Rey, «Evolution of a Ranking Algorithm for Scientific Documents in the Computer Science Area», en *XX Argentine Congress of Computer Science Selected Papers*, La Plata, Buenos Aires, Argentina: EDULP, 2015, pp. 145-155.
- [20] «Modelos de evaluación de producción científica para el área de Ciencias de la Computación». [En línea]. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/45837>. [Accedido: 04-mar-2016].
- [21] R. Adler, J. Ewing, y P. Taylor, «Joint committee on quantitative assessment of research: citation statistics», *Australian Mathematical Society Gazette*, vol. 35, n.º 3, pp. 166-88, 2008.
- [22] J. Bar-Ilan, «Which h-index? — A comparison of WoS, Scopus and Google Scholar», *Scientometrics*, vol. 74, n.º 2, pp. 257-271, nov. 2007.

Marco de trabajo para mejorar el aprovechamiento de factores críticos de éxito en proyectos de ingeniería de software

Acosta Parra, Carlos Alberto¹; Vilallonga, Gabriel^{1,2}; Riesco, Daniel²; Dusso, Juan Pablo³; Zurita Perea, Franco³

¹Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca

Maximio Victoria No 55 - C.P: 4700 - San Fernando del Valle de Catamarca

Telefono: +54(383)443-5112 int. 168

²Departamento de Informática de la Facultad de Ciencias Físico, Matemáticas y Naturales de la Universidad Nacional de San Luis

Ejercito de los Andes 950 - San Luis

Tel: +54(266)452-0300 int. 2127

³Subsecretaría de Tecnologías de la Información del Gobierno de Catamarca

Sarmiento 613 – San Fernando del Valle de Catamarca

Tel: +54(383)443-7500

carlosacostap@tecno.unca.edu.ar, gvilallo@tecno.unca.edu.ar/unsil.edu.ar, driesco@unsil.edu.ar, jduiso@catamarca.gov.ar, fzurita@catamarca.gov.ar

Resumen

A medida que avanzan los años la ingeniería de software ha ido adoptando nuevos enfoques y prácticas mediante los cuales ha logrado evolucionar. ¿Pero esta “evolución” ha repercutido favorablemente en el éxito de los proyectos de ingeniería de software? Diversos estudios demuestran que la tasa de éxitos en proyectos de ingeniería de software todavía sigue siendo baja. Si bien estos nuevos enfoques y prácticas han logrado mejorar la tasa de éxito, su nivel aún continúa fuera de un rango aceptable y confiable. En los diversos estudios realizados se han logrado identificar factores críticos de éxito asociados a los proyectos de ingeniería de software. El presente artículo describe una línea de investigación cuyo principal objetivo es potencializar el

aprovechamiento de esos factores críticos para así incrementar el nivel de éxito de los proyectos de ingeniería de software, a través de un marco de trabajo que podrá ser empleado dentro de las organizaciones.

Palabras clave: Proyectos de ingeniería de software; Factores críticos de éxito; Marco de trabajo; Herramientas de gestión de proyectos de ingeniería de software; Gerencia de la información; Organizaciones; Herramientas para la gerencia de la información.

Contexto

El proyecto de investigación “Fortalecimiento de la Calidad y la Productividad en Ingeniería de Software”, posee una línea de investigación dedicada a la potencialización de factores críticos

de éxito en los proyectos de ingeniería de software. El mencionado proyecto es evaluado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Catamarca y se desarrolla y ejecuta en ámbitos del Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa y en el Departamento de Informática de la Facultad de Ciencias Físico, Matemáticas y Naturales de la Universidad Nacional de San Luis donde los integrantes se desempeñan como docentes de la carrera de Ingeniería en Informática (UNCa), e Ingeniería en Informática y Licenciatura en Ciencias de la Computación (UNSL) respectivamente. A su vez, se cuenta con un convenio doble titulación e intercambio y colaboración docente y de investigación entre la Universidad Nacional de Catamarca y la Universidad de Camerino (Italia). También se cuenta con el convenio establecidos entre la UNSL y la Universidad Federal de Minas Gerais (UFMG), en el marco del programa Centros Asociados para el Fortalecimiento de Posgrados Brasil/Argentina (CAFP-BA) tanto para maestrías como para doctorados. Esta línea de investigación, inserta en el proyecto, establece como prioridad la formación científico-técnico de los integrantes con la premisa de proceder a la transferencia de resultados a los medios informáticos como así también al ámbito científico y tecnológico en general.

Introducción

Los proyectos de ingeniería de software son encabezados por un gestor de proyectos el cual se encarga de su planificación, implementación, monitoreo y control. Es de vital importancia para las organizaciones que los proyectos de ingeniería de software sean exitosos no

solo en aspectos de su ejecución sino también en el valor agregado que aportan a la organización sus resultados. Para ello, es necesario que la organización cuente con el tacto suficiente para tomar decisiones acertadas desde antes de la concepción de sus proyectos de ingeniería de software, para así maximizar las posibilidades de éxitos de los mismos.

El Reporte del Caos ha sido publicado por el Standish Group anualmente desde el año 1994, y se define como una fotografía del estado de la industria del desarrollo de software. Para el reporte del año 2015 (Standish Group International, 2015) han sido estudiados 50000 proyectos alrededor de todo el mundo, contemplando desde proyectos de pequeñas mejoras hasta implementaciones masivas de reingeniería de sistemas.

El Standish Group ha considerado tres causas de fracasos a la hora de analizar los proyectos de ingeniería de software. La primera causa se encuentra relacionada a las demoras en las entregas de los resultados de los proyectos. La segunda causa está vinculada a los problemas relacionados al presupuesto asignado al proyecto. La tercera y última causa proviene de la disconformidad de la organización con los productos resultantes del proyecto. La Figura 2 describe lo mencionada en el presente párrafo.

Los resultados del reporte indican que todavía hay trabajo por hacer en torno a la consecución de los resultados exitosos de los proyectos de ingeniería de software. La Tabla 1 resume los resultados de los proyectos en los últimos cinco años utilizando la nueva definición de factores de éxito (a tiempo, dentro del presupuesto, y con un resultado satisfactorio).

	2011	2012	2013	2014	2015
Exitosos	29%	27%	31%	28%	29%
Apenas aceptables	49%	56%	50%	55%	52%
Fallidos	22%	17%	19%	17%	19%

Tabla 1. Resultados del Reporte del Caos 2015

Una parte fundamental del análisis Standish Group en los últimos 21 años ha sido la identificación y clasificación de los factores que trabajan juntos para hacer que los proyectos resulten exitosos. Los resultados del año 2015 y la clasificación de los factores se muestran en la siguiente tabla.

Factores de éxito	Ponderación
Apoyo de los mandos altos	15
Madurez emocional	15
Participación del usuario	15
Optimización organizacional	15
Equipos cualificados	10
Arquitectura estandarizada	8
Metodologías ágiles	7
Proyectos modestos	6
Un gestor de proyectos con expertise	5
Objetivos de negocio claros	4

Tabla 2. Factores críticos de éxitos en proyectos de ingeniería de software

Líneas de Investigación, Desarrollo e Innovación

Inserto en el marco del proyecto se encuentra la línea de investigación referida a la potencialización de los factores críticos de éxito. Para ello se propone una abstracción desde la perspectiva de los proyectos de ingeniería de software hacia una perspectiva de gerencia de la información dentro de las organizaciones. Por medio de esta abstracción se busca analizar herramientas gerenciales y combinarlas con las herramientas de gestión de proyectos para así lograr la potencialización de los factores críticos que impactan en el incremento de las probabilidades de éxito de los proyectos de ingeniería de software abordados por las organizaciones.

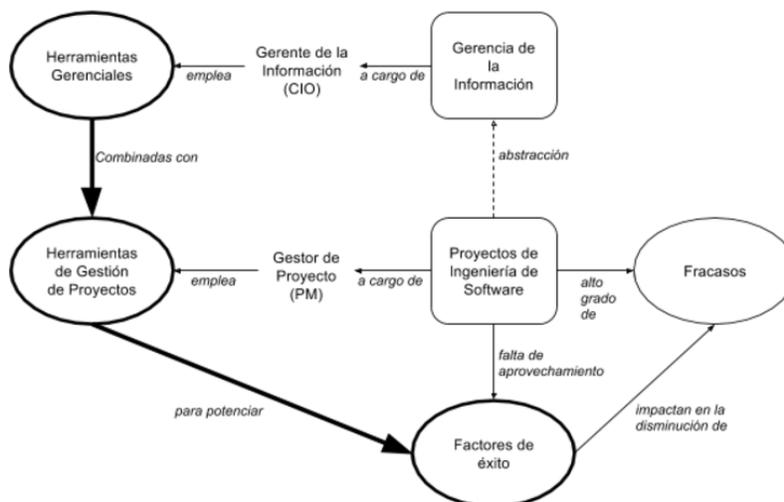


Figura 1. Línea de investigación

¿Cómo potenciar los factores de éxito dentro de los proyectos de ingeniería de software? Es la pregunta en la que se basa esta línea de investigación. Asociada a esta pregunta, esta línea de investigación se encuentra guiada por la siguiente hipótesis: el uso de un marco de trabajo acorde permitirá potenciar los factores de éxito en proyectos de ingeniería de software.

Resultados y Objetivos

De acuerdo a la hipótesis planteada en la línea de investigación, el resultado será un marco de trabajo definido y probado en un ámbito real de aplicación. Es por ello que se ha establecido un vínculo con la Subsecretaría de Tecnologías de la Información del Gobierno de la Provincia de Catamarca. Dicho organismo es el responsable de la definición de políticas y la gestión de las tecnologías de la información de manera transversal a lo largo de todo el Poder Ejecutivo de la Provincia de Catamarca.

En concordancia con lo antes mencionado, se ha planteado como objetivo general del trabajo:

- Elaborar un marco de trabajo para la potencialización de factores de éxito en proyectos de ingeniería de software.

Del objetivo general, se desprenden los siguientes objetivos específicos:

- Indagar acerca de la problemática relacionada a proyectos de ingeniería de software tanto en el Poder Ejecutivo de Catamarca como a nivel mundial.
- Estudiar metodologías, métodos y técnicas relacionadas a los factores críticos de éxito de los proyectos de ingeniería de software.
- Elaborar un marco de trabajo que permita potenciar los factores críticos de éxito en proyectos de ingeniería de software.
- Asistir a la Subsecretaría de Tecnologías de la Información en la realización de una prueba piloto del marco de trabajo elaborado.

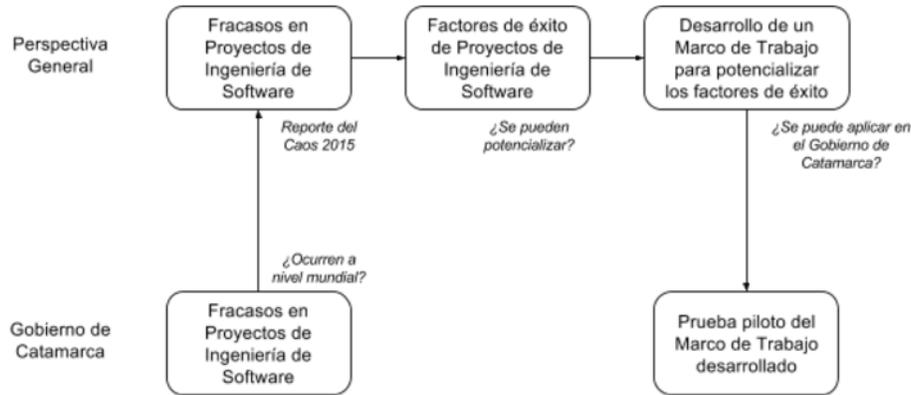


Figura 2. Resumen de la investigación a realizar

Formación de Recursos Humanos

Esta línea de investigación se encuentra en una etapa intermedia, donde un integrante del proyecto se encuentra desarrollando su tesis de maestría en el marco de la carrera de Maestría en Ingeniería de Software de la Universidad Nacional de San Luis en torno al objetivo de esta línea de investigación: el desarrollo de un marco de trabajo para la potencialización de factores críticos de éxito en los proyectos de ingeniería de software. Además los participantes pertenecen a distintas cátedras de las carreras de ingeniería y licenciatura de las universidades que participan. Se ha procedido a la incorporación de alumnos de los últimos años con la finalidad de incorporarlos a actividades de investigación y desarrollo en las áreas específicas del proyecto. El proyecto prevé un programa de capacitación y formación de recursos humanos, que contempla las siguientes actividades:

- Dirección de tesinas de grado de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa integrados al proyecto, como así también en la UNSL.
- Participación de los integrantes del proyecto en cursos de

actualización y posgrado en el área de estudio.

Para garantizar la capacitación y actualización del equipo de investigación, así como la difusión de los avances y resultados logrados, se propuso la participación en eventos nacionales e internacionales de la especialidad, como congresos, simposios, seminarios y cursos.

Referencias

- Ardoino, J. El diseño de procesos en organizaciones efectivas. Edicon, Argentina, 2014.
- Baca Urbina, G. Formulación y evaluación de proyectos informáticos. Mc Graw Hill, México, 2006.
- Davidson Frame, J. La nueva dirección de proyectos: herramientas para una era de cambios rápidos. Granica, Argentina, 2011.
- Heller, M. The CIO Paradox: Battling the Contradictions of IT Leadership. Bibliomotion, USA, 2013.
- Jacobson, I. La esencia de la Ingeniería de Software: Aplicando el Núcleo de Semat. Nueva Librería, Argentina, 2014.
- McCracken, G. Chief Culture Officer: un directivo orientado a la cultura. Océano, México, 2012.
- Standish Group International, The Chaos Report 2015, 2015.

Creación y Simulación de Metodologías de Análisis, Clasificación e Integración de Nuevos Requerimientos a Software Propietario

De Federico Sara¹, Gago Javier¹, Sincosky Noelia¹, Avogradini Mariela¹, Melfi Lucía¹, Moschetti Diana¹, Gonnet Silvio²

¹Área Investigación del Departamento Ingeniería en Sistemas de Investigación
Universidad Tecnológica Nacional Facultad Regional Rosario
Zeballos 134, Rosario, Argentina

²Coordinador. INGAR Instituto de Desarrollo y Diseño, Conicet
Universidad Tecnológica Nacional Facultad Regional Santa Fe
Avellaneda 3657, S3002GJC Santa Fe, Argentina
saraedf@gmail.com, sgonnet@santafe-conicet.gov.ar

Resumen

La priorización de nuevos requerimientos a implementar en un software propietario es un punto fundamental para su mantenimiento, la conservación de la calidad, observación de las reglas de negocio y los estándares de la empresa. Aunque existen herramientas de priorización basadas en técnicas probadas y reconocidas, las mismas requieren una calificación previa de cada requerimiento. Si la empresa cuenta con solicitudes provenientes de varios clientes de un mismo producto, aumentan los factores que afectan a la empresa, las herramientas disponibles no contemplan estos aspectos y hacen mucho más compleja la tarea de calificación. Este trabajo de investigación abarca la realización de un relevamiento de los métodos de priorización y selección de nuevos requerimientos utilizados por empresas de la zona de Rosario, y la definición de una metodología para la selección un nuevo requerimiento, que implica el análisis y evaluación de todas

las implicaciones sobre el producto de software y la empresa, respetando sus reglas de negocio. La metodología creada conduce a la definición de los procesos para la construcción de una herramienta de calificación y priorización de nuevos requerimientos en software propietario que tiene solicitudes de varios clientes al mismo tiempo, con instrumentos de calificación que consideran todos los aspectos relacionados, proveerá técnicas de priorización actuales y emitirá informes personalizados según diferentes perspectivas de la empresa.

Palabras clave: Software, empresas, Rosario, mantenimiento, nuevos requerimientos, metodología, técnicas.

Contexto

El proyecto inicial en donde se comenzaron las tareas llamado “Creación y Simulación de Metodologías de Análisis, Clasificación e Integración de Nuevos Requerimientos a Software Propietario” aprobado por Resolución de

Consejo Directivo N° 680/2012, vigente desde el 1º/01/2013 hasta el 31/12/2015, con código de identificación DISI-INV-2013-001, fue avalado por la Secretaría de Ciencia y Tecnología de la UTN FRRO. Como consecución de este proyecto, se ha elevado a Rectorado el PID “Análisis de lenguajes técnicos y metodologías asociados en el desarrollo de sistemas de priorización de nuevos requerimientos en software propietario”, el mismo se encuentra en proceso de evaluación. En este proyecto se construirá la herramienta de software generada como producto del proyecto anterior, dentro de un entorno ágil, cuya estructura estará disponible para presentación a los alumnos de la carrera, y asimismo trabajando en convenio con la empresa desarrolladora TECSO Ltda, quien brindará sus instalaciones y personal en colaboración con la UTN FRRO.

Introducción

Las empresas desarrolladoras de software se enfrentan a un desafío a la hora de incorporar nuevos requerimientos a un software ya posicionado en el mercado, actualizándose según las nuevas necesidades de los clientes pero conservando sus características, robustez y calidad [1,2]. Además, se han acordado normas que describen al detalle todos los pasos del ciclo de vida del producto, la dinámica entre ellos, la documentación mínima necesaria para un seguimiento completo. Una nueva funcionalidad debe ser agregada minimizando el impacto al sistema, reutilizando objetos y código, y satisfaciendo no sólo la necesidad del cliente que realiza la petición, sino también los otros usuarios del sistema [3], y permitirá afianzar el modelo de madurez del software alcanzado por el producto [4,5]. A su vez, la permanencia de productos de software que han logrado

una posición de privilegio en el mercado, ha dado lugar a la necesidad de investigar sobre cómo mantenerlos, logrando al mismo tiempo captar nuevos clientes y conservar/mejorar el nivel de satisfacción de los clientes actuales.

Las herramientas informáticas de apoyo a la toma de decisiones incorporan técnicas de priorización conocidas que realizan una calificación inicial de cada requerimiento. Esta calificación involucra un conjunto de perspectivas dentro de la empresa que se interrelacionan en forma compleja. En la literatura sobre Ingeniería de Requerimientos se destaca que las técnicas de priorización utilizan comúnmente como criterios para la priorización la importancia, el costo, el daño, la duración, el riesgo, y la volatilidad de los requerimientos [6].

El estudio realizado a las empresas desarrolladoras de software de la ciudad de Rosario y alrededores demostró que la *criticidad* del requerimiento solicitado está estrechamente relacionada al cliente que solicita la nueva funcionalidad [7]. Esto sucede debido a que el comportamiento del cliente solicitante afecta a las diferentes áreas de la empresa en el proceso de desarrollo e implementación del nuevo requerimiento. La historia de los anteriores casos solicitados por un cliente puede determinar la no elección de un nuevo requisito, teniendo en cuenta el comportamiento tanto financiero como personal del mismo, los resultados del desarrollo y su conclusión exitosa o no. Se debe tener en cuenta cómo se manejó el cliente durante la implementación con el personal de soporte o línea de ayuda, y finalmente, la opinión de los demás clientes ante la nueva versión del producto. Además, la línea gerencial puede querer analizar distintos enfoques del futuro de sus productos, por ello es importante contar con mecanismos que

permitan un pronóstico de la incorporación selectiva de nuevos requerimientos.

Líneas de Investigación, Desarrollo e Innovación

El proyecto se dividió en varias etapas, según las tareas a realizar. Estas etapas se fueron adaptando a los resultados de las anteriores, ya que siempre fue prioridad el cumplimiento de los objetivos respondiendo a las necesidades de las empresas. Los objetivos principales fueron los siguientes:

- Recopilación exhaustiva de trabajos y la normativa vigente referente a la etapa de mantenimiento de software en la Ingeniería de Requerimientos
- Estudio de los trabajos y normativas obtenidos para encontrar si existen metodologías de priorización en la selección de requerimientos en la etapa de mantenimiento.
- Creación de criterios de clasificación.
- Creación de métricas de medición del problema.
- Creación de valores e índices de interés para la empresa, para medición del impacto:
- Creación de una de asignación de prioridades a una cola de requerimientos tratando de maximizar valores e índices de interés para la empresa.
- Formulación de los posibles métodos y procedimientos necesarios para la incorporación de una metodología al proceso del mantenimiento del software.

Resultados y Objetivos

Los resultados alcanzados se describen a continuación:

1. El estudio exhaustivo de las normas, estándares y reglamentaciones vigentes sobre mantenimiento de software y priorización de requerimientos muestra zonas oscuras, que no cubren todas las situaciones que surgen en el ciclo de vida del software, sobre todo en la etapa de mantenimiento y versionado[8][9].
2. La recolección de información referente a la toma de decisiones para la selección de nuevos requerimientos permitió definir las variables que se tienen en cuenta en las empresas de la ciudad de Rosario, su valorización y comportamiento e historiales de casos. Se pudo definir que las empresas no utilizan métodos específicos de priorización, sino procedimientos informales y de gran participación del personal jerárquico.
3. Análisis de la información para la extracción de: variables, criterios, escalas, clasificaciones, métricas, métodos y procedimientos definidos y utilizados por las empresas. Se detectaron fallas y ausencias en los procedimientos utilizados que se compensaban con la experticia del personal y el conocimiento informal de los clientes.
4. En base al análisis del punto anterior, se estructuró una metodología para priorización y selección de nuevos requerimientos a implementar durante la etapa de mantenimiento de software.
5. Esta metodología fue revisada para su refinamiento. Con su descripción se crearon diagramas de procesos, que fueron validados. Empresas colaboradoras participaron tanto en el refinamiento como en la validación de los procesos, utilizando casos reales de requerimientos y clientes.

La Figura 1 muestra un extracto de los resultados obtenidos en las primeras etapas del proyecto, en donde se

efectuaron encuestas para la obtención de información sobre los métodos y procedimientos que utilizan las empresas desarrolladoras de software de la ciudad de Rosario para la priorización y selección de nuevos requerimientos.

Punto de importancia en el proceso	Grado de uso	Observaciones
Tiene historial de requerimientos	Muy alto	Historial de pedidos en narrativas
Clasificación de los requerimientos	Muy alto	La escala de valoración es cualitativa, uso relativo de herramientas
Tiene historial de los clientes	Muy alto	Historial financiero no directamente relacionado con cada pedido
Dispersión de la información	Muy alto	A lo largo de todas la áreas
Se genera una nueva versión tras la implementación	Muy alto	Distribución compulsiva a todos los usuarios
Problemas con el desarrollo	Muy alto	Muy variados, íntimamente relacionados con el usuario solicitante
Diferenciación entre bugs y requerimientos	Muy alto	Se tratan por diferentes vías
Interés en poseer una herramienta integral de priorización	Alto	Predisposición a la implementación
Aplica Criterio de criticidad para priorizar	Alto	Definición diferente para cada empresa
Otras formas de clasificación	Alto	Amplia variedad y características
Interés en la opinión del usuario solicitante	Alto	Encuestas de satisfacción y mesas de ayuda al cliente
Seguimiento del desarrollo	Leve	Sólo algunas áreas de la empresa
Uso de métricas y métodos	Leve	Algunas empresas, no se detallan cuáles
Impacto financiero	Leve	Diferentes técnicas para evaluar
Impacto sobre el producto	Leve	Test de producto
Documentación integral	Bajo	Datos dispersos o incompletos
Uso de herramientas informáticas para la priorización	Bajo	Herramientas de monitoreo del ciclo de vida del software
Captación de opinión de los otros usuarios	Bajo	Concentración en el usuario solicitante

Fig 1. Resultados de las encuestas sobre priorización de requerimientos

Luego, a partir de la información obtenida se perfiló una estructura preliminar de metodología. Esta estructura inicial se mostró a un grupo de empresas colaboradoras para su análisis y corrección. Tomando ejemplos reales como prueba, se detectó el seguimiento de todas las posibilidades que pudieran surgir dentro de un proceso de selección de un nuevo requerimiento.

En base a los datos colectados del análisis en las empresas, se corrige la

metodología, contemplando todos los casos e información necesaria para llevar a cabo un proceso de priorización y selección que considere todas las variables que intervienen en la implementación en un software. Se crean estructuras de datos para contener toda la información, que está dispersa a lo largo de toda la empresa desarrolladora, y recabada por diferentes áreas y sectores. La Figura 2 muestra un ejemplo de intento de implementación en una empresa colaboradora, en donde la información que se utiliza para la priorización de un nuevo requerimiento se encuentra en varias áreas y es retenida en diferentes formatos y medios de recopilación, físicos, digitalizados en repositorios distintos e inconexos, etc. Esta información se unifica en una base de datos de Clientes.

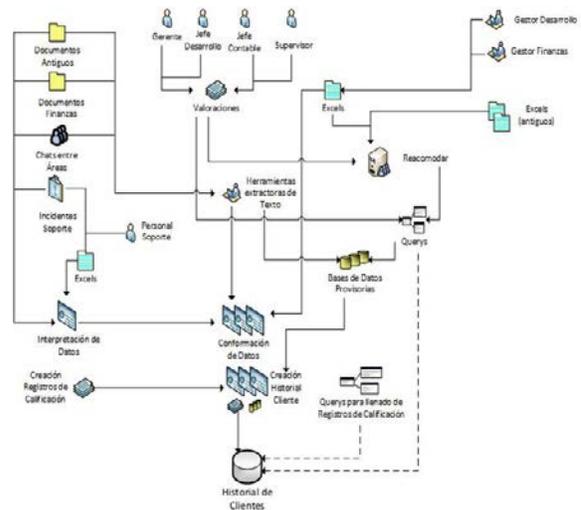


Fig 2. Unión de información de Clientes de empresa desarrolladora de Software

En base a distintas pruebas y análisis en conjunto con las diferentes empresas, usando varios casos reales, observando los puntos críticos y optimizando cada uno de los pasos involucrados, se llegó a la construcción de diagramas de procesos que definen la metodología en forma eficaz y validada[10][11][12]. Estos

diagramas muestran los tres puntos más importantes de este tema, la captación de un nuevo requerimiento, la obtención de información del cliente que lo solicita, y la priorización del mismo. Las Figura 3, 4 y 5 muestran los diagramas de procesos de esta metodología.

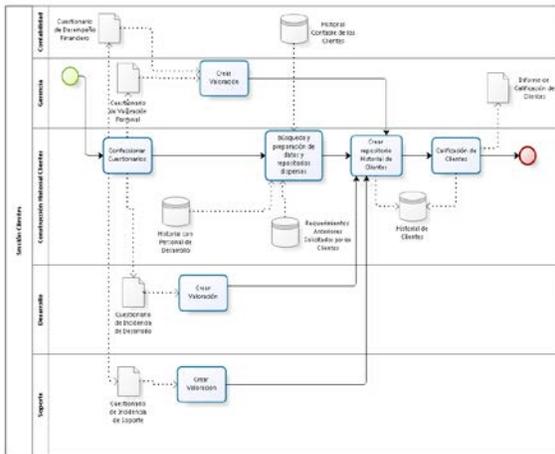


Fig 3. Diagrama de procesos de la unificación de información de Clientes.

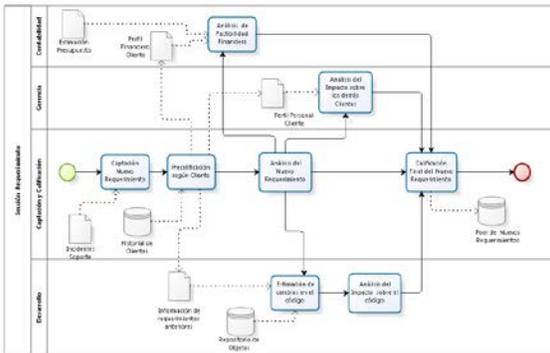


Fig 4. Diagrama del proceso de la captación y clasificación de un nuevo requerimiento.

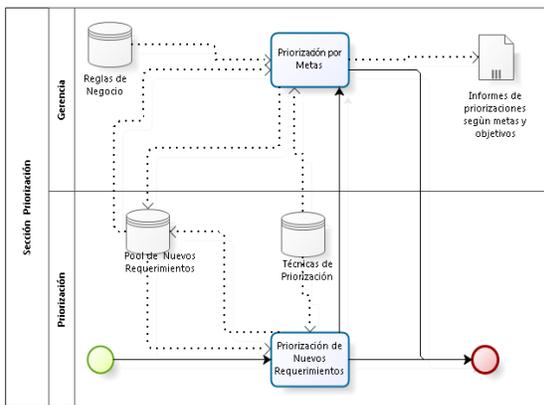


Fig 4. Diagrama del proceso de la priorización de un nuevo requerimiento.

Conclusiones

Gracias a la gran participación de las empresas y al trabajo de todos los integrantes, junto con la coordinación y guía por parte del Dr. Gonnet, este proyecto llegó a la obtención de la definición de una metodología para la priorización de nuevos requerimientos de software en etapa de mantenimiento, la cual permitirá la construcción de una herramienta informática, que facilitará y proveerá los medios para resolver óptimamente el arduo proceso de calificación de los requerimientos.

Formación de Recursos Humanos

El proyecto cuenta con una investigadora categoría IV, docentes de la Facultad Regional Rosario con desempeño profesional en el mercado de software de la ciudad, tres alumnos de grado a punto de recibirse, ya insertos en el mercado laboral de sistemas y capacitándose en investigación, dos estudiantes de la Maestría en Ingeniería en Sistemas de Información. Se han presentado trabajos a varios congresos presentando la evolución de las etapas. Además el proyecto es base de la tesis de maestría de una de las estudiantes. Para el año 2016 se ha presentado un proyecto a nivel Universidad que es continuación de este, en donde se la construcción de una herramienta informática de priorización de nuevos requerimientos en un entorno ágil, en donde alumnos podrán trabajar y relacionarse con empresas desarrolladoras, u observar un proceso ágil de desarrollo.

Referencias

- [1] Canfora G., Cimitile A., "Software Maintenance". *University of Sannio, Faculty of Engineering at Benevento Palazzo Bosco*

Lucarelli, Piazza Roma 82100, Benevento Italy,
29 November, 2000

[2] IEEE 12207-2008, “*Systems and software engineering -- Software life cycle processes*”, <http://standards.ieee.org/findstds/standard/12207-2008.html>

[3] Sher F., Jawawi D., Mohamad R., Babar M., “Multi-Aspects Based Requirements Prioritization Technique for Value-Based Software Developments” *International Conference on Emerging Technologies ICET* (2014)

[4] Erdil, K. et al: “*Software Maintenance as Part of the Software Life Cycle*”, Department of Computer Science Tufts University (2003)

[5] April A., Desharnais J., “*Software Maintenance Maturity Model (SMmm): A Software Maintenance Process Model*”, Department of Software Engineering École de technologie supérieure 1100 Notre-Dame West, Montréal, Québec, Canada. (2006)

[6] Pohl K., “*Requirements Engineering. Fundamentals, Principles, and Techniques*” Springer (2010)

[7] De Federico S., Sincosky N., Lascano A., Avogradini M., Moschetti D. “Análisis para la identificación de clusters en información recopilada de empresas de desarrollo de software sobre técnicas de priorización de requerimientos”, *Actas del 2º Congreso Nacional de Ingeniería Informática/Sistemas de Información 2ª CoNaIISI Vol I*. 2014. Issn: 1171-1176

[8] De Federico S., Sincosky N., Avogradini M., Moschetti D. “Construcción de una Metodología para la Priorización y Selección de Nuevos Requerimientos a Implementar en Software en Etapa de Mantenimiento”. *Actas del WICC 2015, Workshop de Investigadores en Ciencias de Computación. Salta, Argentina*. 2015. ISBN: 978-987-633-134-0.

[9] Achimugu P., Selamat A., Ibrahim R., Mahrin M.: *A systematic literature review of software requirements prioritization research*, *Information and Software Technology*, <http://dx.doi.org/>

[10] Aksyonov K., Spitsina I., Aksyonova O., Schaible P., “The state analysis of intelligent tools of Computer Aided Software Engineering”, *24th International Crimean Conference of*

Microwave & Telecommunication Technology CriMiCo (2014)

[11] Okawa T., Kaminishi T., Hirabayashi S., Koizumi H., Sawamoto J.: An Information System Development Method Based on the Link of Business Process Modeling with Executable UML Modeling and its Evaluation by Prototyping, 22nd International Conference on Advanced Information Networking and Applications - Workshops, AINAW (2008)

[12] Bizagi, *Business Process Management Software*, <http://www.bizagi.com/es/>

Modelado guiado por arquitecturas de software para un sistema de control de tráfico aéreo en territorio Argentino

Gabriela Vilanova¹, Silvia Rivadeneira Molina², Diana Cruz²

Instituto de Tecnología Aplicada (ITA) Dpto Cs Exactas y Naturales

¹Unidad Académica Caleta Olivia

²Unidad Académica Río Turbio

Universidad Nacional de la Patagonia Austral

Acceso Norte Ruta 3 Cp 9011 Caleta Olivia, Santa Cruz, 0297 4854888

e-mails vilanova@uolsinectis.com.ar, sgrivadeneira@yahoo.com.ar, dianacruz@gmail.com

Resumen

El control del tráfico aéreo (ATC) es una de las aplicaciones de software más exigentes. Es una aplicación en tiempo real, indispensable para la seguridad nacional. Vidas humanas se pueden perder si el sistema no funciona correctamente o bien deja de funcionar por un lapso de tiempo, y es altamente distribuída, lo que requiere decenas de controladores trabajando en cooperación para orientar aeronaves a través del sistema de rutas aéreas.

El objetivo del presente trabajo es aplicar un proceso de diseño guiado por la arquitectura para un sistema de gran escala y de gran criticidad como lo es el software de asistencia al controlador aéreo de un Centro de control de área.

Palabras clave: sistemas críticos, atributos de calidad, arquitectura de software.

Contexto

El proyecto de investigación PI 29B176 UNPA Modelado y diseño de software, un enfoque arquitectural, participa del programa de incentivos, se ha iniciado en 2014, es de tipo I, se basa en el proyecto finalizado 29B134 (2012-2014) es de apoyo a las carreras de pre grado, grado y posgrado del área informática que se dictan en UNPA en Unidades Académicas Caleta Olivia, Río turbio y río Gallegos. Dichas carreras son: Licenciatura en sistemas, Ingeniería, Analista de Sistemas y Tecnicaturas en Diseño Web y en Redes. Maestría en sistemas. Los integrantes son docentes, investigadores categorizados de dos unidades Académicas de UNPA a saber Unidad académica Río Turbio (UART) y Unidad académica Caleta Olivia. (UACO) hay un integrante docente con posgrado de la Universidad de Magallanes, docente visitante de UART.

Introducción

La arquitectura de software juega un papel fundamental en el desarrollo a gran escala de software de calidad. Los ingenieros de software generalmente se refieren a la arquitectura de software como “la organización a nivel de un sistema de software integrado de los componentes, las relaciones entre ellos y las restricciones”.[1][2]

Además, la arquitectura de un sistema es responsable de la captura de las abstracciones de la arquitectura esencial para garantizar un conjunto relevante de los factores de calidad. Según la definición de la arquitectura de software que figura en [1], diseño basado en la arquitectura se refiere principalmente a las siguientes cuestiones:

1. La comunicación entre las partes interesadas en el sistema o stakeholders, es decir usuarios u operadores, clientes, diseñadores, programadores etc. La arquitectura de software representa una abstracción común de un sistema pudiéndose utilizar por todos los interesados como base para el entendimiento mutuo, la negociación, el consenso y la comunicación.
2. Principios de las decisiones de diseño, es decir aplicación de heurísticas y patrones o estilos de diseños que se irán ajustando a los requerimientos de calidad deseados.
3. Abstracción transferible de un sistema. La arquitectura de software constituye el modelo intelectualmente comprensible de un sistema, de cómo está estructurado y cómo sus elementos trabajan juntos. Este modelo, es decir las distintas vistas se

puede aplicar a otros sistemas que exhiban atributos de calidad y requisitos funciones similares pudiéndose promover el reúso a gran escala.

Varios estilos arquitectónicos y patrones recurrentes han sido reconocidos y codificados por la comunidad de arquitectura de software. [2,3]. Hay, en consecuencia, cierto acuerdo sobre el conjunto de normas de diseño para identificar qué tipo de componentes y conectores pueden ser usados para componer un sistema dentro de cada estilo, junto con las limitaciones locales o globales sobre la forma en que esta composición se debe hacer. Algunos de estos patrones implican algún tipo de razonamiento para lograr los atributos de calidad [4].

1.1 Proceso de diseño guiado por la Arquitectura de Software.

El diseño de una arquitectura no es una actividad independiente, pero es un paso dentro del desarrollo y proceso evolutivo del producto. Un ciclo de vida típico consiste en cuatro procesos iterativos. El proceso externo se refiere a la evolución de los requerimientos de un producto durante su maduración. El próximo proceso refiere a desarrollo iterativo del producto. Por ejemplo, cuando se construye sistema, el cliente puede no saber especificar exactamente cuáles son los requerimientos que tendrá. Sin embargo, esto puede ser descubierto durante un proceso de desarrollo iterativo. Los dos procesos iterativos siguientes definen los métodos de diseño de arquitectura orientado a los atributos de calidad, la iteración interna donde la arquitectura de software es diseñada, evaluada y transformada por sus requerimientos de calidad, y una iteración

externa donde la selección de requerimientos es realizada.

El foco de la iteración interna es la evaluación y transformación de los atributos de calidad. Los métodos de diseño proveen soporte para un objetivo, el proceso de diseño racional, balanceando y optimizando especialmente los requerimientos de calidad. El método iterativo evalúa el grado en que la arquitectura soporta cada requerimiento de calidad y mejora la arquitectura usando transformaciones hasta que todos los requerimientos de calidad hayan sido cumplidos.

El método de diseño de la arquitectura de software no solo define los procesos asociados con el diseño de la arquitectura, sino además un número de artefactos que documentan la arquitectura y las decisiones de diseño que encabezan una arquitectura.

Estos artefactos son divididos en dos amplias categorías, la especificación de requerimientos y la arquitectura de software. La especificación de requerimientos a su vez se divide en requerimientos funcionales y requerimientos de calidad. La arquitectura de software consiste en cuatro artefactos, el contexto del sistema, los arquetipos, la estructura arquitectural y las decisiones de diseño.

El contexto del sistema define las interfaces del sistema de software con su contexto. Los arquetipos representan las abstracciones centrales que posee el sistema cuando es construido. La estructura representa la descomposición

de la arquitectura en sus componentes centrales y sus relaciones entre esos componentes. Existen tres tipos de decisiones de diseño, transformaciones, restricciones y reglas.

1.2 Modelado basado en Arquitectura de Software. Sistema de control de tráfico aéreo.

El Centro de control de área: (ACC) es la dependencia encargada del control de Área y parte del control de Aproximación. La función principal desempeñada por el ACC es brindar seguridad, ordenamiento y rapidez al tránsito aéreo en su jurisdicción. El ACC tiene jurisdicción sobre una FIR (Región de Información de Vuelo), controla las aeronaves desde que éstas liberan la zona controlada por las TWR hasta la siguiente zona de control o hasta que ingresa a la FIR controlada por otro ACC. Los vuelos están siempre bajo el control de una dependencia determinada, a excepción de aquellos que se realizan fuera de los espacios aéreos controlados, tales como vuelo deportivo o fumigadores. Por ejemplo el ACC Comodoro Rivadavia (ACC CR) comprende toda la zona sur del país, desde Bahía Blanca hasta Usuahia e islas Malvinas (Ver Figura 1).



Figura 1. Carta de navegación en rutas aéreas (Argentina Sur)

El ACC CR tiene comunicación vía radio con todos los aeropuertos ubicados bajo su área de control y brinda los permisos de control de tránsito aéreo para todas las aeronaves que por allí circulan, incluso los cruces del territorio argentino en tránsito hacia otros países. Como se mencionó al principio, el ACC realiza parte del control de aproximación, que es cuando las aeronaves abandonan el nivel de crucero para descender hacia el aeropuerto de destino. Las tareas del ACC son múltiples y complejas, por lo que cuenta con mayor cantidad de personal que una TWR. Las distintas divisiones del espacio aéreo realizadas dentro del ACC se denominan *Circuitos*, y cada uno de éstos es controlado por un controlador habilitado y un ayudante. No se debe confundir ACC con TWR, ésta se encarga de los arribos y despegues, en cambio el ACC controla principalmente la ruta.

PNL (Plan de Vuelo): es la Información especificada que, respecto a un vuelo proyectado, o a parte de un vuelo de una aeronave, se somete a las dependencias de los servicios de tránsito aéreo, concretamente, el PLN contiene información acerca de un vuelo determinado indicando origen, destino, nivel de vuelo en ruta, tiempo de vuelo, aeropuertos de alternativa, autonomía, etc; es presentado a la autoridad aeronáutica por el piloto o su representante (despachante de aeronaves)

en la oficina ARO-AIS. Cuando un vuelo progresa desde su aeropuerto de salida a su aeropuerto de llegada, se ocupan varias entidades ATC que guía de forma segura a través de cada porción de rutas aéreas (y las instalaciones de tierra Figura 3.) que está utilizando.

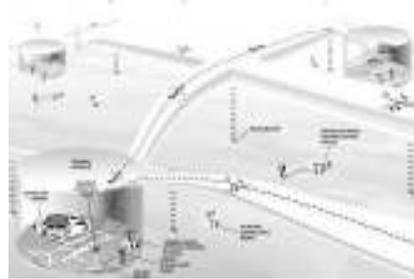


Fig. 2 Volar desde el punto A al punto B en el sistema de control del tráfico aéreo

En resumen, el sistema del ATC debe hacer lo siguiente:

- Convertir las hojas de ruta y los ingresos que realice el controlador sobre las comunicaciones con los pilotos y los controladores de aeropuertos dentro de su jurisdicción, Cada consola elige los informes que necesita para mostrar, cualquier consola es capaz de mostrar cualquier zona en la pantalla como si en tiempo real tuviéramos un radar disponible (Simulación Dinámica).

- Manejar alertas de conflicto, el sistema deberá transmitir al controlador las potenciales colisiones de aviones o cualquier eventualidad que pueda resultar en riesgo para las aeronaves.

- Proveer capacidad de respaldo de información si un evento de falla sucede.

- Proporcionar la posibilidad de grabación de las comunicaciones para su posterior reproducción.

- Estar disponible en todo momento

Líneas de Investigación, Desarrollo e Innovación

Las distintas líneas de trabajo pretenden cubrir en la temática de modelado y diseño de sistemas críticos, complejos, áreas de interés de estudio tales como modelado de requerimientos funcionales y no funcionales desde punto de vista de arquitecturas de software, arquitecturas orientadas a servicios aplicando procesos ágiles, de manera de disminuir el impacto de decisiones tomadas de diseño (patrones de grano fino y grueso) logrando niveles de calidad aceptables tanto en producto como en proceso de desarrollo del software.

Resultados y Objetivos

Se ha definido el proceso arquitectural, como empezar a pensar una arquitectura, como lograr que en distintos niveles cada uno de los atributos de calidad sean satisfechos por la arquitectura lograda. El sistema de apoyo deberá tener una disponibilidad total y un alto rendimiento alto debido a su larga vida prevista, su gran tamaño y su importante función dentro de su campo de aplicación. Para ello la arquitectura propuesta debe utilizar una amplia gama de mecanismos de tolerancia a fallos, incluyendo la redundancia en hardware y software, y un proceso de detección de fallas por capas para que el fallo no se propague dentro del sistema.

Formación de Recursos Humanos

El proyecto de investigación PI 29B176 se ha iniciado en 2014, es de tipo I, se basa en proyecto finalizado 29B134 (2012-2014) es de apoyo a las carreras de informática que se dictan en UNPA, Licenciatura en sistemas, Ingeniería, Analista de Sistemas y Tecnicaturas en Diseño Web y en Redes y Maestría en sistemas. Los integrantes son docentes, investigadores categorizados de dos unidades Académicas de UNPA a saber UART y UACO. La temática ha sido abordada como caso de estudio para trabajo final de cátedra en asignaturas de carreras analista de sistemas y optativa Arquitectura de software de Ingeniería en sistemas.

Referencias

- [1] Bass, P. Clements and R. Kazman. *Software Architecture in Practice*. Addison Wesley.
- [2] D. Garlan and M. Shaw. An Introduction to software architecture. In *Advances in Software Engineering and Knowledge Engineering*, pages 1-39, Singapore, 1993. World Scientific Publishing Company.
- [3] Bosch Jan Design and Use of Software Architectures. Addison Wesley.
- [4] B. Boehm, P. Bose, E. Horowitz and M. J. Lee. Software requirements negotiation and renegotiation aids: A theory-W based spiral approach. In *Proc 17th International Conference on Software Engineering*, 1994.

Las TIC al Servicios del Dato Abierto: Situación Actual, Conceptualización e Iniciativas de Apertura de Información Pública

Carolina I. Chayle, Claudia M. Herrera, María A. Barrera

Departamento de Informática/Facultad de Tecnología y Ciencias

Aplicadas/Universidad Nacional de Catamarca

Maximio Victoria N° 55 - C.P: 4700 - San Fernando del Valle de Catamarca

Telefono: 03833- 435112 – int 168

cchayle@gmail.com, cm_herrera@tecno.unca.edu.ar, mbarrera@tecno.unca.edu.ar

Resumen

El Estado recopila y produce grandes volúmenes de datos, tales como datos de tipo climatológico, energía, economía, salud, medio ambiente, agricultura, defensa, seguridad pública, social, cultural, presupuestos, entre muchos otros. En general el ciudadano desconoce la existencia de estos datos, que pueden tener accesos restringidos, o estar en formatos no estandarizados o privados, lo que lleva a preguntarse ¿cómo hacer para que los ciudadanos puedan sacar el máximo provecho de estos datos?; ¿qué tipo de servicios debieran ofrecer los gobiernos para aumentar la participación ciudadana en las iniciativas de apertura de la información pública? En la actual sociedad, donde la información es un activo fundamental y estratégico, se está haciendo cada día más significativo un movimiento a favor de la publicación de Datos Abiertos (DA), para ser utilizados de manera directa por cualquier usuario. Se pretende realizar una evaluación general sobre la situación actual de las TIC al servicio del dato abierto contextualizando el análisis de los datos en iniciativas de apertura de información pública en Argentina y en particular en la

región; produciendo una contribución efectiva, de aportes concretos que impliquen nuevas propuestas TIC que asistan al desarrollo de software específico.

Palabras clave: TIC, Dato Abierto, Gobierno Abierto, Sector Público.

Contexto

Se pretende en este proyecto establecer métodos que abordarán el estudio de los conceptos fundamentales sobre DA, su importancia y las diversas metodologías, técnicas, herramientas y propuestas que se encuentran disponibles para evaluar y diagnosticar su implementación. Se trabajara en los siguientes aspectos de DA en el sector Publico tanto de la Argentina como de la Región, como ser Eficiencia en la gestión de las ciudades, Sector infomediario emergente, Surgimiento de nuevos empleos, Vehículo para implantar la transparencia, Participación y colaboración ciudadana y áreas de aplicación de la reutilización.

1

Introducción

En los últimos años ha comenzado una auténtica revolución en materia de acceso a la información pública por parte de los ciudadanos. A esta revolución se le ha denominado “Open Data” (OD) o “Datos Abiertos” (DA), que consiste en poner a disposición de la sociedad los datos de interés común de la ciudadanía para que, de cualquier forma, éstos puedan desarrollar una nueva idea o aplicación que entregue nuevos datos, conocimientos u otros servicios que el gobierno no entrega.

Hoy en día, gracias a las posibilidades tecnológicas ofrecidas por Internet y por la Web, y por las potencialidades de hardware y software existentes, se está observando a nivel mundial el surgimiento de una tendencia que reconoce en éstas y otras herramientas de las Tecnologías de la información y las Comunicaciones (TIC), pasando por Big Data, cloud computing y redes sociales, la nueva frontera de la gobernanza política, de la relación entre ciudadanía y gobierno y de la gestión pública en su amplio sentido. Asimismo, cuando se habla de reutilización de los DA en las Administraciones Públicas, se debe analizar también la importancia del DA en el marco de las ciudades inteligentes (open smart cities).

El modelo de datos abiertos asigna un rol especial a los intermediarios de datos, desarrolladores de software, investigadores o periodistas que trabajan con datos y otros que tienen las habilidades necesarias para transformar datos en información y aplicaciones que beneficien a los ciudadanos comunes. De hecho, la ciudadanía necesita de aplicaciones y servicios que conviertan los datos en productos consumibles y que aporten valor en su vida cotidiana. Por eso, las iniciativas de datos abiertos deben contemplarse como una oportunidad para el estímulo de la innovación.

Los siguientes puntos describen algunos de los aspectos clave de los DA en el sector público:

- *Eficiencia en la gestión de las ciudades:* el primer reutilizador de los datos de DA suele ser la propia administración que lo implanta, que encuentra ahí un repositorio de información fiable y fácilmente procesable.
- *Sector infomediario emergente:* incluye al conjunto de empresas que generan productos y/o servicios para su comercialización a terceros, a partir de la información del sector público.
- *Surgimiento de nuevos empleos:* la apertura de datos fomenta la innovación y la creación y/o adaptación de empresas en torno a servicios y tecnologías de internet que pueden generar negocio a partir de la reutilización de datos.
- *Vehículo para implantar la transparencia:* La adopción de estrategias de apertura y reutilización de DA y de gobierno abierto (Open Government), permiten una mayor participación de cualquier persona en el diseño y la puesta en marcha de servicios públicos, dando a la ciudadanía el poder para tomar decisiones más inteligentes en su vida diaria.
- *Participación y colaboración ciudadana:* Con la política de apertura de datos crecen los contenidos que pueden hacerse públicos en las webs locales. Asimismo, los conjuntos de datos expuestos permiten que la ciudadanía los procese para generar transparencia.
- *Múltiples áreas de aplicación de la reutilización:* Información de datos geográficos, información geoespacial, medioambiental, meteorológica, recursos turísticos,

transporte, tráfico, e incidencias de tráfico en tiempo real, resultados electorales, viviendas alquiladas, atlas de mortalidad, localización de centros sanitarios, farmacias, hospitales, centros de salud, locales comerciales, registros de asociaciones y fundaciones, datos de contrataciones administrativas, ayudas y subvenciones, trámites, registros, entre otros son algunos ejemplos de las áreas donde se pueden reutilizar los DA.

Como afirman Ramírez-Alujas y Dassen (2014), “Estamos convencidos que el gran avance de la tecnología, en particular de las TICs, son, y serán, el mecanismo preferido a adoptar por aquellas administraciones públicas que deseen transformar su relación con la ciudadanía, como así también aspectos organizacionales internos”

Líneas de investigación, desarrollo e Innovación

Las plataformas de publicación de DA constituyen un elemento fundamental como soporte tecnológico de cualquier estrategia, por ello se deberán analizar las plataformas con mayor difusión y uso a nivel nacional, haciendo especial hincapié en el grado de compatibilidad que ofrecen con los estándares existentes tanto en nuestro país como en el resto del mundo, para lograr la selección más adecuada de la tecnología que permita la consecución de los objetivos de una iniciativa de apertura de la información pública. En este aspecto, también será necesario realizar un análisis pormenorizado de las características ideales que deberían estar presentes en cualquiera plataforma, tanto de elaboración propia como en un producto disponible en el mercado, ya sea mediante un modelo de software gratuito o comercial.

Es necesario destacar que el medio será el principal proveedor de casos de estudios que servirán como casos reales para la prueba de técnicas y metodologías en el chequeo de su eficiencia, de esta manera se podrán realizar críticas centradas en sistemas concretos de aplicación. Los convenios realizados con empresas de desarrollo de software local, como así también organizaciones gubernamentales, que se han rubricado en colaboración mutua con la Universidad Nacional de Catamarca, brinda los recursos necesarios para cumplimentar con la propuesta de trabajo.

Resultados y Objetivos

La presente investigación pretende analizar la situación actual de iniciativas de DA en Argentina y la región, sobre su utilización por parte de la sociedad y de la capacidad de los organismos de ofrecer información relevante y de calidad, con el objeto de facilitar el acceso de la ciudadanía; modernizar y hacer más eficiente y transparente el funcionamiento del Estado; permitir y alentar la reutilización de datos para la creación de aplicaciones y servicios derivados para brindar información primaria y accesible. Para ello se llevarán a cabo acciones tales como:

- Estudiar y analizar el rol de las TIC como mediadoras y facilitadoras del acceso al DA.
- Determinar las características ideales que deberían estar presentes en cualquiera plataforma, tanto de elaboración propia como en un producto disponible en el mercado, ya sea mediante un modelo de software gratuito o comercial.
- Analizar plataformas para la selección de una tecnología que permita la obtención de los objetivos de una

iniciativa de apertura de la información pública.

- Identificar herramientas y estrategias de apertura y reutilización de datos públicos (DA).
- Conocer cómo colaboran los distintos organismos, públicos y privados argentinos, para garantizar el acceso público y gratuito a la información.

La mayor parte de los integrantes del equipo son docentes de la carrera Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca (UNCA) y al momento de la presentación de este proyecto se encuentran en la etapa de elaboración de sus tesis de posgrado en áreas relacionadas directamente a la temática del proyecto, lo que permite asegurar la capacidad técnica del grupo para el estudio propuesto. Asimismo, el grupo cuenta ya con algunos trabajos relacionados al área, presentados en reuniones científicas y en la revista de Producción Científica de la Facultad de Tecnología y Ciencias Aplicadas de la UNCA.

Formación de Recursos Humanos

Para garantizar la capacitación y actualización del equipo de investigación y como forma de ratificar y evaluar los avances y resultados obtenidos en la investigación, se propone la participación en eventos científicos regionales, nacionales e internacionales de la especialidad, como congresos, simposios, seminarios y cursos, por ello el programa de capacitación y formación de recursos humanos, contempla las siguientes actividades:

- Participación en cursos de actualización y posgrado en el área de estudio.
- Incorporación de alumnos de los últimos años de la carrera de

Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCA., en calidad de becarios y auxiliares de investigación.

- Participación de los integrantes del proyecto en la dirección, asesoramiento y evaluación de tesinas de grado de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCA.
- Desarrollo de tesis de posgrados por parte de los docentes integrantes.
- Proponer un programa de capacitación y transferencia de tecnologías relacionadas a la publicación de DA al Gobierno Local y a las industria del software local.

En cuanto a la formación específica del equipo de docentes investigadores, se destaca que los mismos se encuentran abocados a actividades de capacitación y estudios de posgrado, para la elaboración de 3 (tres) tesis correspondiente a la carrera Maestría en Ingeniería del Software de la Universidad Nacional de San Luis, y 2 (dos) tesis Maestría en Informática de la Universidad Santo Tomás de Aquino.

Referencias

CONCHA G. and NASER A. *Datos abiertos: Un nuevo desafío para los gobiernos de la región*. Instituto Latinoamericano y del Caribe de Planificación Económica y Social (ILPES). 2012.

OUI. *Factores Técnicos y Económicos que Facilitan o Inhiben la Implementación de Datos Abiertos en la Región - Guía Implantación*. Organización Universitaria Interamericana (OUI). 2012.

Conferencia Regional de Datos Abiertos para América Latina y el Caribe. *Datos gubernamentales en formato abierto?*

SOLAR M. and CONCHA G. *Modelo de Madurez Open Data*.

UNLAR. *Open Data. Miradas y Perspectivas de los Datos Abiertos*. 2015. Conferencia Regional de Datos Abiertos para América Latina y el Caribe. *¿Qué son los Datos Abiertos (Open Data)?*

RAMÍREZ-ALUJAS A. and DASSEN N. *Vientos de Cambio. El avance de las políticas de Gobierno Abierto en América Latina y el Caribe*. Banco Interamericano de Desarrollo. 2014.

Aplicación de principios SOA para la composición de servicios en ambientes ubicuos.

Germán Montejano^{1,2}; Oscar Testa²; Rubén Pizarro²; Darío Segovia²; Oscar Dieste³; Efraín R. Fonseca C.⁴

¹Facultad de Ciencias Físico-Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950 – (5700) San Luis – San Luis – Argentina
Tel.: +54-2652-424027 – Int. 251

gmonte@unsl.edu.ar – web: <http://www.unsl.edu.ar>

²Facultad de Ciencias Exactas y Naturales
Universidad Nacional de La Pampa
Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina
Tel.: +54-2954-425166– Int. 28

[otesta, ruben]@exactas.unlpam.edu.ar

³Facultad de Informática
Universidad Politécnica de Madrid
Campus Montegancedo – (28660) Boadilla del Monte – Madrid – España
Tel.: +34 91 336 5011

odieste@fi.upm.es

⁴Departamento de Ciencias de la Computación
Universidad de las Fuerzas Armadas ESPE
erfonseca@espe.edu.ec

Resumen

Con la aparición de las redes de computadoras y la conectividad que nos ofrecen los dispositivos móviles actuales, se conforma un ambiente propicio para satisfacer las necesidades de interacción entre distintos actores de la vida cotidiana. Estos dispositivos poseen sensores de distinta índole, facilidad de conectividad, incluso en áreas con poca señal o acceso a redes. La composición de servicios en ambientes ubicuos debe resolver los problemas planteados por el contexto, la heterogeneidad y las contingencias de los dispositivos, así como también la personalización de los mismos. Debido a que los dispositivos en donde los servicios son ejecutados, poseen limitaciones de recursos, se deben tomar en cuenta consideraciones especiales acerca de la eficiencia y rendimiento de la composición de servicios. En este trabajo presentamos una propuesta de solución de coordinación de servicios en ambientes ubicuos utilizando

las características y especificaciones estándares de la arquitectura orientada a servicios (SOA). Se hacen aportes a la tecnología de servicios, de ambientes pervasivos y dispositivos ubicuos.

Palabras clave: sistemas ubicuos, SOA, servicios, composición de servicios.

Contexto

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de Software, Conceptos, Métodos y Herramientas en un Contexto de “Ingeniería de Software en Evolución” – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis y en el Proyecto de Investigación: Ingeniería de Software: composición de servicios en ambientes ubicuos – Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa. Las líneas aquí presentadas actualmente forman parte de las bases de un anteproyecto de tesis doctoral y de una tesis de maestría, ambas en ingeniería de

software por la Universidad Nacional de San Luis.

Introducción

Con la aparición de las redes de computadoras y la conectividad que nos ofrecen los dispositivos móviles actuales, se conforma un ambiente propicio para satisfacer las necesidades de interacción entre distintos actores de la vida cotidiana. La tendencia actual es hacia los ambientes ubicuos, los cuales se caracterizan por estar poblados de numerosos sensores que gracias a la integración extrema de los elementos electrónicos, son invisibles al usuario y están en permanente rastreo de la actividad humana [8].

Podemos indicar que la computación ubicua es un desarrollo tecnológico que intenta que las computadoras no se perciban en un entorno como objetos diferenciados, y que la utilización por parte de los seres humanos sea lo más transparente y cómoda posible, facilitando de esta manera la integración en la vida cotidiana.

Dispositivos ubicuos son todos aquellos dispositivos que pueden existir en todas partes, es decir, son dispositivos electrónicos que tienen capacidad de procesamiento y comunicación y pueden ser encontrados en lugares diversos de la vida cotidiana, cuya particularidad es que poseen escaso tamaño.

Desde hace varios años los dispositivos ubicuos han ganado importancia y presencia en la vida cotidiana de las personas, debido principalmente a que: poseen distintos tipos de sensores (posicionamiento, proximidad, luminosidad, temperatura, etc.), facilitan la conectividad incluso en áreas con poca señal o acceso a las redes, permiten la convergencia tecnológica (computo, medios, telefonía, etc) y

brindan acceso a servicios de distinta índole (mapas, ayudas, etc).

Todas estas características han permitido que por ejemplo, la cantidad de móviles existentes en el mercado se aproxime a la cantidad de habitantes mundiales, según un informe de la Unión Internacional de Comunicaciones (UIT, por sus siglas en inglés). En este informe, se estima que hasta finales del año 2014 hubieron casi 7.000 millones de suscripciones de telefonía celular, lo que corresponde al 96% de la población global; es decir, que la cantidad de usuarios de telefonía móvil se acercara al número de personas que viven en el planeta [6].

La proliferación de dispositivos computacionales ubicuos e interconectados, así como los avances de las comunicaciones entre dispositivos, ha permitido que estos sean generadores y consumidores de servicios al mismo tiempo, es decir que de acuerdo a las capacidades del dispositivo puede ofrecer a otros equipos sus funciones y así formar conectividad entre ellos.

La composición de servicios presenta la forma en que se pueden combinar o enlazar servicios disponibles para llevar adelante una tarea determinada. En ambientes ubicuos, la composición de servicios, presenta nuevos desafíos tales como: la heterogeneidad, las contingencias de los dispositivos y la personalización de los mismos (por ej. provisión de servicios de acuerdo a las preferencias del usuario). Dado que los dispositivos en donde los servicios son ejecutados poseen limitaciones de recursos (ej. poca memoria y batería), se deben hacer consideraciones especiales respecto a la eficiencia y rendimiento de la composición de servicios [5].

Basados en el estado de las distintas tecnologías de composición de servicios y coordinación en sistemas ubicuos, tanto

de forma teórica como práctica, propondremos un modelo de coordinación y orquestación de dispositivos ubicuos en ambientes pervasivos, en el cual nos encontramos trabajando actualmente. Este modelo pretende atacar la problemática de la composición de servicios en estos ambientes basándose en las tecnologías existentes y ya probadas de SOA.

Líneas de Investigación, Desarrollo e Innovación

Actualmente las composiciones y coordinaciones en sistemas ubicuos se necesitan hacer con elementos de un mismo proveedor, para que los protocolos de comunicación entre ellos sea el que provee la empresa que los comercializa, es decir, con protocolos y herramientas de programación privativos. También se puede avisorar la complejidad que implica coordinar estos dispositivos para que en conjunto puedan realizar la tarea programada.

Las distintas soluciones que se encuentran implementadas hoy en día están faltas de estandarización tanto en la conectividad de los distintos dispositivos, así como de los protocolos que permiten la comunicación entre ellos. Existen ejemplos de dispositivos que cuentan con la característica de NFC (Near Field Communication), pero que solamente pueden ser implementados y conectados con dispositivos del mismo fabricante, de la misma manera que sucede hoy en día con las balizas inteligentes, o como es el caso de la domótica, donde la estandarización se encuentra todavía lejos de poder ser utilizada en proyectos reales. Nuestra propuesta es poder utilizar y aplicar todas las ventajas, especificaciones y estandarizaciones

existentes en SOA para la coordinación de servicios disponibles en ambientes pervasivos a través de la utilización de dispositivos ubicuos. Si pensamos que cada dispositivo ubicuo en un ambiente pervasivo es proveedor o consumidor de un servicio encaja perfectamente en la arquitectura de servicios. La diferencia con las tecnologías actuales es que la comunicación y coordinación entre ellos (hasta hoy solamente disponible en equipos con ciertas características de procesamiento) no se realiza únicamente por protocolos de TCP/IP, ya que la conectividad entre ellos puede ser de distinta índole o manera (bluetooth, RFID, contacto, etc). Es decir que si nosotros pudiéramos estar en una capa por encima de la comunicación (tal como lo hacen las capas del modelo de red OSI) no estamos muy distanciados de la posibilidad de reutilización de las tecnologías existentes.

Resultados y Objetivos

Por lo expuesto, vemos que existe un campo de trabajo importante en el desarrollo de composición de servicios en ambientes ubicuos, más precisamente en la coreografía de servicios, la cual no es abordada en los estudios previos de la materia.

De esta manera el objetivo de la investigación es la realización de un modelo de composición de servicios -más precisamente coreografía de servicios- en ambientes pervasivos y con dispositivos ubicuos, que permita la realización de coreografías de una manera dinámica, autónoma y confiable.

Este modelo está basado en estándares abiertos y existentes como los que se han visto en este trabajo y basándose en la arquitectura de servicios (SOA), así como también en los protocolos y lenguajes de composiciones ya existentes.

Hasta el momento, en la investigación hemos realizado el análisis del estado del arte, que ha dado las bases para poder enfocar claramente el resto de las etapas de la investigación.

Se trabajó en la realización de un prototipo, basado en la arquitectura planteada Figura 1, que permite la evaluación del trabajo realizado, y una aproximación al modelo final que se espera elaborar y en el cual nos encontramos trabajando actualmente. La complejidad en este punto de desarrollo se centra básicamente en los protocolos de comunicación entre los distintos dispositivos, ya que no se basará únicamente en los estándares de TCP/IP-Internet. En cada dispositivo se deberá evaluar qué alternativas de comunicación posee el mismo y se adaptará el modelo propuesto a las posibilidades de conexión existente. Se espera que el manejo de mensajes entre los dispositivos esté en una capa superior a la de comunicación, tomando en este caso como protocolo base de mensajes los existentes hoy en día en las especificaciones de SOAP y REST.

Para poder simular los dispositivos y las interacciones entre los mismos, se está trabajando con dispositivos móviles con tecnología NFC y sistema operativo Android¹, con dos plataformas de bajo costo Raspberry Pi B+, con una placa Arduino Mega, además de tarjetas y etiquetas RFID. Se cuenta además para el prototipo con sensores de movimiento y ultrasónicos, los cuales pueden ser conectados a la plataforma Raspberry Pi o bien a la placa Arduino. También se cuenta con un módulo Bluetooth para dar conectividad de red a la placa Arduino Mega, lo que permite simular una manera

¹La selección de este sistema operativo es debido a que ya se cuenta con estos elementos con sensores acordes a las necesidades y a que se cuenta con experiencia previa en el desarrollo de aplicaciones en dicho sistema operativo.

de comunicación del dispositivo. La plataforma Raspberry cuenta con una placa wi-fi para dar conectividad de red de forma inalámbrica.

En la etapa de la especificación del modelo, se espera poder formalizar el modelo respectivo por el cual se puedan llevar adelante coreografías en ambientes pervasivos con dispositivos ubicuos. El modelo final a construir tendrá en cuenta también las distintas limitantes existentes tanto en los dispositivos y el ambiente de desarrollo y aplicación, ya que debemos recordar que los mismos son pervasivos o ubicuos, lo que hace que la disponibilidad de los mismos sea, o mejor dicho, pueda ser limitada, así como su capacidad de procesamiento y mantenimiento de estados a lo largo de la coreografía.

Podemos, entonces, concluir que la composición y coordinación de dispositivos ubicuos en ambientes pervasivos está sometido a varios desafíos. Por otro lado hemos visto que la arquitectura SOA posee la potencialidad necesaria para responder a los desafíos que mencionamos, con el agregado de que nos brinda la ventaja de tratarse de una tecnología conocida y afianzada en el ambiente de sistemas distribuidos, con estandarizaciones de mucha índole y en los campos más importantes de la materia y que a su vez cuenta con una amplia base de lenguajes, frameworks y herramientas disponibles en el mercado, de utilización directa.

Nuestra investigación, a través del objetivo planteado, permite trasladar progresivamente los conceptos de SOA a dispositivos ubicuos en ambientes pervasivos, permitiendo además conocer y estudiar las ventajas e inconvenientes de la aplicación de la arquitectura de servicios en dichos ambientes.

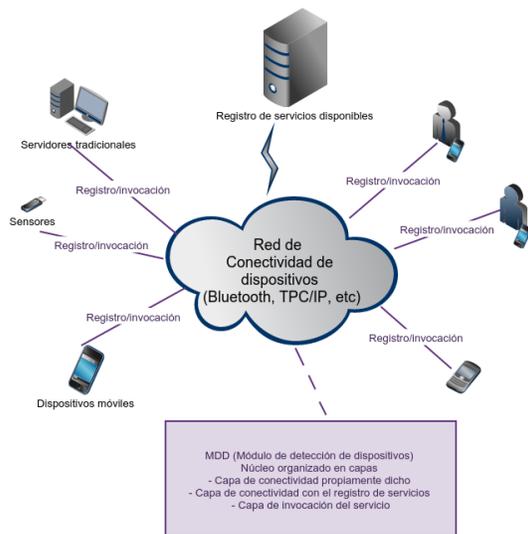


Figura 1. Arquitectura propuesta
Formación de Recursos Humanos

Además de los resultados obtenidos/esperados en el punto 3, se espera como resultado en la formación de recursos humanos, la continuación de esta misma línea de proyecto como tesis doctoral de alguno(s) de los investigadores. También se espera lograr una mayor interrelación con la Universidad de Minas Gerais con la que se cuenta con un convenio con tal objetivo como parte de él. Se espera avanzar también en un convenio de colaboración con la Universidad Politécnica de Madrid para la aplicación de las metodologías aquí presentadas en los proyectos de Ingeniería de Software Empírica. Adicionalmente, se espera que otras tesis de Maestría, así como tesinas de Licenciatura surjan a partir de los logros obtenidos en la presente línea de investigación.

Referencias

[1] S. Najar, M. K. Pinheiro, and C. Souveyet. A New Approach for Service Discovery and Prediction on Pervasive Information System.

Procedia Computer Science, 32:421–428, 2014. The 5th International Conference on Ambient Systems, Networks and Technologies (ANT-2014), the 4th International Conference on Sustainable Energy Information Technology (SEIT-2014).

[2] O. f. A. S. I. S. (OASIS). Web Services Coordination (WS-Coordination). <http://docs.oasis-open.org/ws-tx/wstx-wsat1.1-spec-os/wstx-wsat-1.1-spec-os.html>, 02 2009.

[3] F. Palmieri. Scalable service discovery in ubiquitous and pervasive computing architectures: A percolation-driven approach. *Future Generation Computer Systems*, 29(3):693–703, 2013. Special Section: Recent Developments in High Performance Computing and Security.

[4] M. L. f. C. Science and M. A. I. Laboratory. Pervasive Human-Centered Computing. Available at <http://oxygen.csail.mit.edu/>, Last visited: Aug 15th,2002.

[5] Q. Z. Sheng, X. Qiao, A. V. Vasilakos, C. Szabo, S. Bourne, and X. Xu. Web services composition: A decade’s overview. *Information Sciences*, 280(0):218–238, 2014.

[6] U. I. T. (UIT). Unir on Internacional de Telecomunicaciones. https://www.itu.int/net/pressoffice/press_releases/2014/23es.aspx, 10 2015.

[7] M. Viroli. On competitive self-composition in pervasive services. *Science of Computer Programming*, 78(5):556–568, 2013. Special section: Principles and Practice of Programming in Java 2009/2010 & Special section: SelfOrganizing Coordination.

[8] M. Weiser. Hot topics-ubiquitous computing. *Computer*, 26(10):71–72, Oct 1993.

[9] Testa, Oscar; Riesco, Daniel; Montejano, Germán; Debnath, Narayan. “Rigorous Definition in RAISE Specification Language of a Framework for Web Services about Geographic Information Systems”. 7th International Conference on Information Technology: New Generations. Abril 12-14, 2010, Las Vegas, Nevada, USA.

[10] Debnath, Narayan.; Testa, Oscar; Riesco, Daniel; Montejano, Germán. “Geographic Information Systems: RSL Services Definitions of a Framework for Web Services”. CATA-2011: ISCA 26th International Conference on Computers and their applications. March 23-25, 2011, New Orleans, Louisiana, USA.

[11] Testa, Oscar; Riesco, Daniel; Montejano, Germán. “Especificación Formal en RSL de una Infraestructura abierta y estándar de Servicios Web para Sistemas de Información Geográfica”. CACIC – 2009.

Aplicaciones Móviles: arquitecturas, visualización, realidad aumentada, herramientas de medición, desarrollo híbrido

Marta C. Fennema, Susana I. Herrera, Rosa A. Palavecino, Pablo J. Najjar Ruiz, Paola D. Budán, Gabriela I. Suárez, Melisa Córdoba

Instituto de Investigación en Informática y Sistemas de Información, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero
{rosypgg, sherrera}@unse.edu.ar, najarpablo@yahoo.com.ar, {pbudan1, suarezgabriela03}@gmail.com, cba_melisanv@yahoo.com.ar

Resumen

Se presenta el estado de avance de una investigación sobre sistemas móviles, iniciada en el año 2012 [11, 17, 20] y que culminará en este año 2016 (fue prorrogado un año desde Diciembre de 2015).

Las aplicaciones móviles poseen desventajas originadas en restricciones en el tamaño de pantalla, en la capacidad de procesamiento y en la disponibilidad de energía. Ante ello, este proyecto estudia e investiga acerca de métodos, técnicas y herramientas que permitan optimizar la calidad de los sistemas móviles, tomando como referencia el estándar de calidad de la norma ISO/IEC 25000 y la norma de procesos de ciclo de vida del software ISO/IEC 12207.

Al inicio del proyecto, los investigadores pertenecían a diferentes universidades nacionales del NOA; sin embargo, en la actualidad, la mayoría pertenecen a la Universidad Nacional de Santiago del Estero.

Las principales tecnologías que se estudian son: arquitecturas móviles alternativas, herramientas para el desarrollo híbrido de aplicaciones móviles, técnicas de visualización, realidad aumentada, nuevos métodos ágiles de desarrollo aplicables a sistemas móviles, herramientas para la medición de magnitudes físicas, sensibilidad al contexto.

Las aplicaciones o prototipos que se desarrollan se aplican en los siguientes dominios: educación, turismo, ingeniería forestal, rehabilitación (salud).

Palabras clave: Sistemas móviles, calidad de sistemas móviles, arquitecturas de diseño, aplicaciones híbridas, realidad aumentada, visualización.

1 Contexto

Esta investigación se lleva a cabo en el marco del proyecto denominado “Optimización de la calidad de los Sistemas Móviles mediante la implementación de nuevas arquitecturas, realidad aumentada, técnicas de visualización y redes móviles Ad-Hoc. Aplicaciones en m-learning y en gestión del conocimiento”. A su vez, el

proyecto forma parte de un Programa de Investigación en el área Informática que tiene como objetivo realizar investigación aplicada sobre nuevas tecnologías que permitan mejorar la calidad tanto del proceso como del producto software.

Actualmente, el equipo de investigación del proyecto está conformado por docentes del Instituto de Investigaciones en Informática de la Facultad de Ciencias Exactas de la Universidad Nacional de Santiago del Estero (UNSE) y del Departamento de Computación de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta (UNSa).

Además, el proyecto cuenta con el asesoramiento de investigadores de Institutos de Investigación de la Facultad de Informática de la Universidad Nacional de La Plata.

Está financiado por el Consejo de Ciencia y Técnica de la Universidad Nacional de Santiago del Estero, durante el período Enero 2012-Diciembre 2016.

2 Introducción

Los sistemas móviles son los sistemas que se desarrollan para que sean ejecutados desde dispositivos móviles, usando redes de telefonía celular y redes wi-fi. Se ejecutan desde diversos sistemas operativos, donde se destaca Android, dado que es el más usado en el mercado argentino. Los sistemas móviles pueden ser ejecutados en cualquier momento y en cualquier lugar, una de sus principales ventajas es la ubicuidad, que proviene del pequeño tamaño del dispositivo. Pero esto último trae como consecuencia importantes restricciones de recursos: poca capacidad de procesamiento, escasa memoria, tamaño pequeño de pantalla, entre otros aspectos [3, 29, 34, 39]. Por ello, el desarrollo de sistemas móviles involucra cuestiones propias e importantes, como ser: el diseño de interacción, posicionamiento, representación del espacio, formas de censado, tratamiento del contexto, arquitecturas y herramientas de implementación [10, 27]. Estas características constituyen las variables que impactan directamente en la calidad de la aplicación que se desarrolla [2, 23, 30, 41].

En cuanto a la arquitectura de diseño, un sistema móvil generalmente se basa en una aplicación cliente que se conecta a un servidor de aplicaciones alojado en la nube [5, 8, 19, 39]. Este servidor, a su vez, utiliza los servicios de un proveedor de ubicación, un Sistema de Información Geográfico (GIS) y de la información provista por diversos puntos de interés. Sin embargo existen aplicaciones que se ejecutan totalmente en el cliente [36]. Las arquitecturas alternativas son tres: Arquitectura Servidor, Arquitectura Cliente, Arquitectura Cliente-Servidor o Híbrida. En esta investigación se desarrollan aplicaciones con las tres arquitecturas. En cuanto a la arquitectura híbrida y cliente, se trabajan principalmente con el sistema operativo Android [9, 25].

Cuando se desconoce la posición del usuario, se la obtiene utilizando una de las siguientes técnicas de sensado de posiciones: GPS, sistema de antenas, tags [8]. Para el posicionamiento in-door es más apropiado el bluetooth o sistema de sensores. En esta investigación se han desarrollado sistemas basados en posicionamiento, utilizando códigos QR como tags. Se está estudiando la aplicación de los otros tipos de posicionamiento, que permiten sensibilidad a la ubicación.

Asimismo, en el marco del proyecto se han iniciado investigaciones para estudiar el comportamiento de los sistemas móviles que utilizan el hardware del teléfono para realizar mediciones [37, 15]. Esto con el propósito de aplicar al campo de la industria forestal.

Los teléfonos móviles o celulares utilizan, típicamente, tecnologías de red especialmente desarrolladas para ese tipo de dispositivos las cuales se han ido clasificando en diferentes generaciones [1, 39]. En Argentina, las empresas de telefonía celular han empezado a ofrecer el servicio 4G recién a fines del año 2014. Actualmente, en la mayoría del país, sigue vigente la 3^o Generación (3G), caracterizada por la convergencia de voz, datos y acceso inalámbrico a Internet a mayor velocidad; es apta para aplicaciones multimedia (envío y recepción de imágenes estáticas y video, servicios de ubicación geográfica, televisión en tiempo real, juegos, etc.). Recién desde mediados de 2015 los usuarios empezaron a migrar masivamente sus dispositivos móviles 3G a 4G.

Las características propias de este tipo de sistemas hacen que sea necesario seguir métodos apropiados para su desarrollo. En el ámbito académico y de la industria existen diversos estudios realizados sobre métodos alternativos de desarrollo [4, 31, 33, 33, 40], entre los que se destacan la metodología Mobile-D [31]. Es una metodología desarrollada por investigadores del Centro de Investigación Técnica de Finlandia, que habría obtenido una certificación CMMI (Capacidad de Madurez del Modelo de Integración) de nivel 2. Se considera necesario el estudio de las diversas metodologías propuestas y diseñar una metodología propia que considere: las características de estos sistemas, el estándar ISO/IEC 12207 [24], el lenguaje de modelado UML y las buenas prácticas como las

provenientes del Modelo Vista Controlador (MVC) [5, 7, 8, 26, 38].

Además, en este último tiempo se ha planteado la necesidad de estudiar el rendimiento de herramientas que permiten producir rápidamente aplicaciones que corren bajo diferentes sistemas operativos, por ejemplo PhoneGap. Se estudiará la calidad de los productos obtenidos con estas herramientas, así como también beneficios y desventajas desde el punto de vista del equipo de desarrollo.

También es importante dentro del desarrollo de aplicaciones atender sensibilidad al contexto referida a otros aspectos diferentes de la localización. En este sentido, en el proyecto se están desarrollando prototipos de aplicaciones sensibles al tiempo, con el propósito de mejorar la usabilidad de la aplicación en usuarios que requieren rehabilitación.

Con el propósito de mejorar la capacidad de comprensión, se estudian técnicas de visualización. Estas permiten mostrar, con mayor eficiencia, una gran cantidad de variables científicas en la pantalla, mejorando la comprensión.

Cabe resaltar que para el estudio de la calidad se tiene en cuenta el estándar ISO/IEC 25000. Es así como la investigación se dirige a optimizar características tanto del modelo de calidad en uso como del modelo de calidad del producto (eficiencia, usabilidad, accesibilidad, comprensibilidad, compatibilidad, etc.). La optimización se logra mediante la incorporación de: arquitecturas de diseño apropiadas, tecnologías de realidad aumentada y de visualización, marcos de compatibilidad para dominios específicos como el turismo, entre otros elementos.

3 Líneas de investigación y desarrollo

La principal línea de investigación de esta propuesta consiste en el estudio de métodos, técnicas y herramientas que permitan la optimización de los sistemas móviles; lo cual significa aumentar su calidad, tomando principalmente las características de eficiencia, usabilidad, compatibilidad y comprensibilidad [23, 30].

Respecto a la eficiencia (velocidad y administración de recursos), se investigó sobre la relación entre ésta y el diseño e implementación de arquitecturas alternativas para las aplicaciones móviles.

En relación a la comprensibilidad y usabilidad (capacidad de ser aprendido y operado, capacidad de atracción), se investiga sobre la construcción de interfaces de usuario usando técnicas de visualización y de realidad aumentada. Así como también el uso de herramientas como brújula y acelerómetro (con sus librerías correspondientes) que permitan realizar de forma fácil cálculos complejos como la medición de los árboles de un monte.

Otro aspecto que se estudia es cómo desarrollar aplicaciones móviles sensibles al contexto, considerando la variable tiempo.

En relación al enfoque de calidad interna, se estudian métodos ágiles usados o factibles de ser usados en el desarrollo de aplicaciones móviles. Y también dentro de este contexto, se estudiarán las herramientas como PhoneGap.

Todos los estudios realizados serán implementados en la optimización de la usabilidad de sistemas móviles de turismo, rehabilitación (salud), aprendizaje, ingeniería forestal y gestión del conocimiento.

4 Objetivos y resultados

El objetivo general que guía esta investigación aplicada es:

Contribuir al mejoramiento de la calidad de las aplicaciones móviles mediante nuevas arquitecturas de diseño, técnicas de visualización, recursos de realidad aumentada y métodos ágiles de desarrollo.

Los objetivos específicos son:

- a) Definir un ecosistema móvil que identifique y caracterice los diversos componentes involucrados en la Computación Móvil (dispositivos, sistemas operativos, herramientas y librerías de programación y emulación, etc.) de la región NOA, con un nivel de abstracción que permita aplicarlo alternativamente a otros contextos.
- b) Analizar las diversas arquitecturas de aplicaciones móviles (servidor, cliente, cliente-servidor, basados en web-services, etc.) y proponer una arquitectura que aumente su eficiencia en términos de velocidad, uso de memoria, procesador y energía.
- c) Determinar criterios y estrategias que permitan diseñar y construir interfaces de usuario que aumenten la usabilidad y capacidad de comprensión de las aplicaciones móviles, mediante el estudio de técnicas de visualización de información, realidad aumentada, metáforas de interacción, sensibilidad al contexto.
- d) Diseñar una propuesta metodológica que guíe el desarrollo de aplicaciones móviles, considerando los aspectos de usabilidad, eficiencia y comprensibilidad.

En el desarrollo de la investigación se agregó el siguiente objetivo:

- e) Determinar las tecnologías de medición de distancia (componentes electrónicos, interfaces, bibliotecas de funciones) que permitan definir con mayor precisión magnitudes físicas, usando dispositivos móviles.

Los resultados parciales obtenidos hasta el momento son:

- Se logró definir un ecosistema del NOA que se revisa permanentemente debido al avance constante de la tecnología móvil y a la ampliación de la muestra [21, 22]. El ecosistema fue publicado y expuesto en distintos eventos científicos [18, 6].
- Se diseñó e implementó una arquitectura híbrida cliente-servidor sobre la cual se monta una aplicación m-turismo. La arquitectura fue publicada en [18,19]. Sobre ella se hicieron estudios de eficiencia, logrando importantes conclusiones [28].
- También con arquitectura híbrida se desarrolló una aplicación para m-learning: Educ-Mobile, presentada en [13, 14], en Android. Se utilizó MVC.
- Con arquitectura web se desarrolló una aplicación para m-learning: Ima-Colab (aún no publicado). El lenguaje principal utilizado fue PHP y JavaScript. Para el diseño se usó el framework Twitter Bootstrap, con diseño responsivo. Para editar código utilizó Sublime Text 3.
- Se diseñó e implementó una arquitectura cliente sobre la cual se monta una aplicación móvil de ayuda a personas con deficiencia visual que se publicó en [35, 36]. La aplicación desarrollada se llama “Lazarillo Virtual”.
- Respecto a la usabilidad de aplicaciones móviles se ha diseñado un marco para optimizar la usabilidad en el área de Turismo [18]. Se avanzó en estudios de usabilidad [12] en sistemas web, usando criterios que pueden ser aplicables a sistemas móviles. Se estudiaron técnicas de visualización en sistemas web, cuyos resultados serán utilizados en sistemas móviles; estas técnicas se aplicaron a estudios de Hidrogeología. Los resultados parciales con avances en usabilidad fueron comunicados en [6, 18].
- Se inició investigación sobre el uso eficiente de las herramientas provistas en los dispositivos móviles (brújula, acelerómetro, etc.) que sirven para mediciones de objetos físicos. Se publicaron resultados parciales en [15].

En síntesis, los principales avances se obtuvieron en el estudio de eficiencia de arquitecturas alternativas de diseño de sistemas móviles.

Con respecto al objetivo d) actualmente se realizó la revisión del estado del arte en metodologías de desarrollo de aplicaciones móviles de manera de asegurar la calidad, basándose principalmente en el estudio de los métodos ágiles. Se investigará sobre los beneficios y desventajas del desarrollo híbrido utilizando herramientas para la generación automática de código multiplataforma (Android, IOS, etc.).

También se continuará investigando sobre el uso óptimo de herramientas provistas en dispositivos móviles, como brújula y acelerómetro, para utilizarlos en procesos de medición de magnitudes de un bosque; en colaboración con el Instituto de la Llanura Chaqueña de la UNSE.

5 Formación de recursos humanos

La Directora del proyecto pertenece a la Universidad del Norte Santo Tomás de Aquino. La Codirectora, al Instituto de Investigación en Informática y Sistemas de Información de la UNSE. La Asesora en Sistemas Móviles pertenece al Laboratorio de Investigación y Formación en Informática Avanzada (LIFIA) de la UNLP y posee vasta experiencia y conocimientos en el desarrollo de aplicaciones móviles [26, 27].

Esta investigación contribuye al desarrollo de una tesis doctoral en Ciencias Informáticas (UNLP) referida a un marco para el análisis y evaluación de experiencias de m-learning [21, 22] y desarrollo de aplicaciones de m-learning, como así también, al desarrollo de una Tesis de Maestría (UNSE) referida la Modernización en el Poder Judicial de Santiago del Estero.

En el año 2013 se terminó un trabajo final de Ingeniería en Informática (UNCa) sobre aplicaciones móviles para personas con deficiencia visual y dos trabajos finales de Licenciatura en Sistemas de Información (UNSE) sobre arquitecturas y eficiencia de aplicaciones móviles y sobre sistemas de visualización.

Actualmente, se están realizando trabajos finales de LSI sobre: realidad aumentada, desarrollo híbrido de aplicaciones, métodos de desarrollo de sistemas móviles, sistemas sensibles al contexto temporal, herramientas para medición de magnitudes físicas. Se tiene previsto incorporar más alumnos de grado en los próximos años, a efectos de brindarles un marco apropiado para la elaboración de sus trabajos finales.

Referencias

1. Aravind, A., Tahir, H. *Towards modeling realistic mobility for performance evaluations in MANET*; University of Northern British Columbia, Prince George, BC, Canada; 2010.
2. Ardagna, C.A., Jajodia, S. *Privacy Preservation over Untrusted Mobile Networks*. CSIS-George Mason University; 2007.
3. Bartz, J. *Mobile Computing Deployment and Management: Real World Skills for CompTIA Mobility+ Certification and Beyond*. Ed. Sybex (Wiley Brand). ISBN: 978-1-118-82461-0. Indianapolis, Indiana, 2015.
4. Blanco, P., Camarero, J., Fumero, A., Werterski, A., Rodríguez, P. *Metodología de desarrollo ágil para sistemas móviles. Introducción al desarrollo con Android y el iPhone*. Tesis de Doctorado en Ingeniería de Sistemas Telemáticos. Universidad Politécnica de Madrid. 2009.
5. Burbeck, S. *Applications Programming in Smalltalk-80(TM): How to use Model-View-Controller (MVC)*. (TM) Smalltalk-80. 1997. Disponible en <[\[www.cs.uiuc.edu/users/smarch/st-docs/mvc.html\]\(http://www.cs.uiuc.edu/users/smarch/st-docs/mvc.html\)>. Consultado el 01/03/2013.](http://st-</div><div data-bbox=)

6. Campos, M. M., Morales, C., Thir, J. M., Herrera, S., *Sistema WEB de Información Hidrogeológica Basado en el Modelo Unificado de Visualización*. I. IX Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del Noa. ISSN 1853-7871. Santiago del Estero, Octubre 2013.
7. Deacon, J. *Model-View-Controller (MVC) Architecture*. Última revisión: 2009. Disponible en: <<http://www.jdl.co.uk/briefings/mvc.pdf>>. Consultado el 01/03/2013.
8. Denso Wave. *Qrcode.com*. Sitio oficial del Estándar QR Code. Disponible en: <http://www.qrcode.com/en/index.html>>. Consultado el 01/03/2013.
9. Emmanouilidis, C., Koutsiamanis, R.A., Tasidou, A. *Mobile guides: Taxonomy of architectures, context awareness, technologies and applications*. Journal of Network and Computer Applications. Volume 36, Issue 1, Pages 103–125. Elsevier, January 2013.
10. Fortier, A., Rossi G., Gordillo S., Challiol, C. *Dealing with Variability in Context-Aware Mobile Software*. Journal of Systems and Software 83(6): 915-936 (2010).
11. Fennema, M., Palavecino, R., Herrera, S. y Najjar Ruiz, P. *Métodos, Técnicas y Herramientas para Optimizar la Calidad de los Sistemas Móviles*. XVII Workshop de Investigadores en Ciencias de la Computación (WICC). ISBN 978-987-633-134-0. Salta. 2015.
12. Gallo, F. R., Palavecino, R. A., Herrera, S. I. *Evaluación de la Usabilidad en sistemas E-Cultura*. 2° Congreso Nacional de Ingeniería Informática/Sistemas de Información. ISSN 2346-9927. San Luis, Argentina. Noviembre, 2014.
13. Herrera, S. I., Najjar, P., Morales, M.I., Sanz, C. y Fennema, M. C. *Educ-Mobile. Juego educativo colaborativo para m-learning*. DEMO. Congreso TE&ET. ISBN 978-987-28186-0-9. Chilecito, La Rioja. 2014.
14. Herrera, S. I. y Sanz, C. *Collaborative m-learning practice using Educ-Mobile*. International Conference on Collaboration Technologies and Systems (CTS). Ed. IEEE. Pp 363 – 370. ISBN: 978-1-4799-5157-4. Minneapolis, MN, USA. 2014.
15. Herrera, S. I., López, G., Ríos, M. *Nuevas Tecnologías Móviles para la Gestión del Conocimiento*. Primer Congreso Internacional Gran Chaco Americano (CONGRACHA 2014). Santiago del Estero. Noviembre, 2014.
16. Herrera, S. I., Gallo F., Najjar Ruiz, P. J. *Accesibilidad en Aplicaciones Móviles para turismo*. Segundo Congreso Argentino de Interacción Persona Computadora, Telecomunicaciones, Informática e Información Científica. ISBN 978.88.96.471.25.8. Córdoba, Nov. 2013.

17. Herrera, S. I., Najar Ruiz, P. J., Rocabado, S. H., Fennema, M. C., Cianferoni, M. C., Optimización de la Calidad de los Sistemas Móviles. Workshop en Investigación en Ciencias de la Computación, WICC. ISBN 978-987-28179-6-1. Paraná 2013.
18. Herrera, S. Najar, P. Contreras, N. Fennema, C., Lara, C., Marco Sistémico para el Desarrollo de Aplicaciones de m-Turismo. IX Congreso Brasileiro de Sistemas. Palmas, Tocantis, Brasil. ISBN 978-85-89102-43-8. Octubre 2013.
19. Herrera, S. I., Najar Ruiz, P. J., Ledesma, E., Rocabado, S. Sistema de Información Móvil para Turismo Receptivo. Revista Gestao e Conhecimento, Edición Especial, Anales del 8° Congresso Brasileiro de Sistemas. ISSN 1808-6594. Pozo de Caldas, Octubre 2012.
20. Herrera, S. I., Fennema, M. C., Rocabado, S., Goñi, J., Optimización de la calidad de los Sistemas Móviles mediante la implementación de nuevas arquitecturas, realidad aumentada, técnicas de visualización y redes móviles Ad-Hoc. L. Workshop en Investigación en Ciencias de la Computación, WICC. ISBN 978-950-766-082-5. Posadas, Abril 2012.
21. Herrera, S. I. & M. C. Fennema. Tecnologías Móviles Aplicadas a la Educación Superior. Congreso Argentino de Ciencias de la Computación. La Plata, 2011.
22. Herrera, S. I., J. L. Goñi & M. C. Fennema. El m-learning en la educación universitaria de posgrado. Jornadas de Ingeniería del NOA. Catamarca, 2011.
23. International Standar Organization. ISO/IEC 25000. Estándares de Calidad del Software. 2011.
24. International Standar Organization. Norma ISO/IEC 12207. Procesos de Ciclo de Vida del Software. 2011.
25. Komatineni, S., MacLean, D. Pro Android 4. Ed. Apres. ISBN 1430239301, 9781430239307. 2012.
26. Lliteras, A., Challiol, C., Gordillo S. Juegos Educativos Móviles Basados en Posicionamiento: Una Guía para su Conceptualización. 13th Argentine Symposium on Software Engineering, 41 JAIIO, La Plata, 2012.
27. Lliteras, A., Challiol, C., Mostaccio, C., Gordillo S. Representaciones enriquecidas para la navegación indoor-outdoor en aplicaciones móviles. XVII Congreso Argentino de Ciencias de la Computación. La Plata, 2011.
28. Najar, P., Ledesma, E., Rocabado, S., Herrera, S., Palavecino, R. Eficiencia de aplicaciones móviles según su arquitectura. CACIC 2014. ISBN 978-987-3806-05-6. La Matanza, Buenos Aires. Octubre, 2014.
29. Pernici, B. Mobile Information Systems. Infrastructure and Design for Adaptivity and Flexibility. Ed. Springer-Verlag, Germany, 2006.
30. Piattini M., García F. y Caballero I. Calidad de Sistemas Informáticos. AlfaOmega, México, 2007.
31. Pikkarainen, M. Mapping Agile Software Development on to ISO 12207. Information Technology for European Advancement. Febrero, 2006. Disponible en: <http://www.agile-itea.org/public/deliverables/ITEA-AGILE-D2.9_v1.0.pdf>. Consultado el 01/03/2013.
32. Rahimian, V. Performance evaluation of mobile software systems: Challenges for a software engineer. 5th International Conference on Electrical Engineering, Computing Science and Automatic Control CCE. Noviembre, 2008.
33. Rahimian, V., Ramsin, R. Designing an Agile Methodology for Mobile Software Development: A Hybrid Method Engineering Approach. The Second IEEE International Conference on Research Challenges in Information Science (RCIS). Junio, 2008.
34. Reza B'Far. Mobile Computing Principles. Cambridge University Press, 2005.
35. Richard, P., *Sistema Guia Para Personas Con Deficiencia Visual*. XIX Congreso Argentino de Ciencias de la Computación (CACIC 2013). ISBN 978-987-23963-1-. Mar del Plata, Octubre 2013.
36. Richard, P., *Sistema Guia Para Personas Con Deficiencia Visual*. XIX Congreso Argentino de Ciencias de la Computación (CACIC 2013). ISBN 978-987-23963-1-. Mar del Plata, Octubre 2013.
37. Rodríguez, R., Vera, P., Martínez, M., Verbel de la Cruz, L., Vallés, F. Construcción de aplicaciones móviles con acceso al hardware de los dispositivos. XVII Workshop de Investigadores en Ciencias de la Computación (WICC). ISBN 978-987-633-134-0. Salta. 2015.
38. Schiller, J. H., Voisard, A. Location-Based Services. Ed. Elsevier. ISBN 0080491723, 9780080491721. 2004.
39. Talukder, A.K., Ahmed, H., Yavagal, R. Mobile Computing: Technology, Applications, and Service Creation. 2° Edición. McGraw-Hill communications engineering series. ISBN 9780070144576. 2010.
40. Vainio, A. M., Tuunanen, T., Abrahamsson, P. Developing Software Products for Mobile Markets: Need for Rethinking Development Models and Practice. In Proc. of the 38th Hawaii International Conference on System Sciences (HICSS'05). Enero, 2005.
41. Zhang, L., B. Tiwana, Z. Qian. Accurate Online Power Estimation and Automatic Battery Behavior Based Power Model Generation for Smartphones. ACM, 2010.

Extendiendo la meta-arquitectura aportada por el enfoque MDA con conocimiento del dominio

Luis Roqué Fourcade, Liliana Arakaki, Daniel Riesco, Germán Montejano
Facultad de Ciencias Físico Matemáticas y Naturales - Universidad Nacional de San Luis
Ejército de Los Andes 950, San Luis, Argentina, +54 (0266) 4520300

email: araroq@unsl.edu.ar, liliana.arakaki@yahoo.com, driesco@unsl.edu.ar, gmonte@unsl.edu.ar

Resumen

Los métodos y técnicas provistos por la Ingeniería de Software hasta la actualidad, proveen un amplio soporte a la actividad de especificación de prácticamente todos los artefactos producidos por el proceso de Diseño de Software y, en contrapartida, un escaso soporte a la actividad característica y recurrente del proceso de Diseño de Software de síntesis y evaluación de alternativas para la toma de decisiones dejándola, paradójicamente para un proceso ingenieril, librada casi por completo al grado de experticia del desarrollador.

Presentamos aquí una línea de investigación que, a partir de resultados obtenidos en trabajos previos [1], se propone desarrollar una meta-arquitectura como una especialización de la aportada por el enfoque Model Driven Architecture (MDA) [2]. La línea propone realizar dicha especialización extendiendo esta meta-arquitectura con conocimiento de la definición de Diseño de Software aportada por trabajos previos [1]. La intención consiste en anticipar en ésta un conjunto de decisiones de diseño e incluir restricciones que acoten el universo de evaluación para otras, con el objetivo de soportar la actividad de síntesis y evaluación de alternativas para la toma de decisiones, mejorando sensiblemente el perfil ingenieril de la actividad de Diseño de Software.

Palabras clave: Ingeniería de software - Diseño – Meta-arquitectura - UML – Modelado - Transformación - MDA

Contexto

Esta línea de investigación, se desarrolla en el marco del proyecto 'Ingeniería de software: aspectos de alta sensibilidad en el ejercicio de la profesión de ingeniero de software', de la Facultad de Ciencias Físico-Matemáticas y Naturales de la Universidad Nacional de San Luis, bajo el número 22F222.

Introducción

Los avances incorporados por los métodos y técnicas actuales son realmente significativos. En el área de especificaciones, mediante lenguajes y estándares que soportan la especificación de todos los

artefactos de software involucrados en el proceso de desarrollo de software y, en el área de proceso, mediante la identificación y consideración de las actividades requeridas para construir dichas especificaciones. Sin embargo, en el área de proceso, se requieren avances aún más significativos. Esto es así porque estas especificaciones, describen solo aspectos estáticos de las actividades y adolecen de soportes más concretos a la ejecución propiamente dicha de la actividad. Esta situación se ve reflejada en múltiples aspectos de estos métodos y técnicas. En particular, la actividad característica y recurrente de síntesis y evaluación de alternativas para la toma de decisiones queda librada casi por completo al grado de experticia del desarrollador.

En la dirección de modelar y soportar el proceso de desarrollo de software, se destacan los esfuerzos y avances realizados por Object Management Group (OMG) [3] y, entre éstos, el enfoque MDA [2]. Son muchos y de gran importancia los aportes realizados por este enfoque, pero aquí estamos interesados en particular en el modelo del proceso de desarrollo de software que propone. Define una arquitectura para el proceso y un modelo de transformaciones constituyendo una meta-arquitectura a partir de la cual es posible obtener la arquitectura de múltiples ocurrencias del proceso de desarrollo de software. La gran importancia de este aporte es clara, sin embargo, adolece aún de un grado de abstracción que dificulta (o imposibilita) la derivación a partir de esta meta-arquitectura de soportes efectivos para actividades concretas de desarrollo de software como lo es por ejemplo la actividad de síntesis y evaluación de alternativas para la toma de decisiones que aquí nos ocupa.

Por este motivo y siguiendo con el enfoque propuesto en trabajos presentados previamente [1] [4] [5], proponemos aquí especializar esta meta-arquitectura extendiéndola con resultados de estos trabajos de manera tal que el proceso de instanciación de procesos de desarrollo de software sea capaz de derivar una arquitectura más especializada en el proceso de Diseño de Software,

en particular en la actividad de síntesis y evaluación de alternativas para la toma de decisiones. Para este fin, desarrollamos a continuación, los conceptos que fundamentan la propuesta y luego presentamos la hipótesis que fundamenta la línea de investigación sugerida.

Model Driven Architecture (MDA) [2]

Object Management Group (OMG) [3], es un consorcio internacional de estándares de tecnología, fundado en 1989, sin fines de lucro y de membresía abierta. Los estándares impulsados por OMG, que incluyen Unified Modeling Language (UML) [6] y Model-Driven Architecture (MDA) [2], hacen posible entornos poderosos de diseño visual, ejecución y mantenimiento de procesos de software y otras clases de procesos [3].

El enfoque MDA permite derivar valor de modelos y arquitectura para soportar el ciclo de vida completo de sistemas físicos, organizacionales y de tecnología informática, con el objetivo de ayudar a manejar complejidad e interdependencia en sistemas complejos. Apoyándose en la definición de conceptos como sistema, modelo, metamodelo, arquitectura y capa arquitectural, y en el hecho de que todos son modelos, estructura y ordena el proceso de desarrollo definiendo dos arquitecturas basadas en niveles de abstracción, desde puntos de vista diferentes pero ortogonales entre sí [7].

Arquitectura de Meta-modelado

El “meta-modelado” es clave en MDA. El enfoque MDA define para esta actividad una arquitectura basada en niveles de abstracción donde cada nivel define un nivel de meta-modelado más concreto que el superior y más abstracto que el inferior.

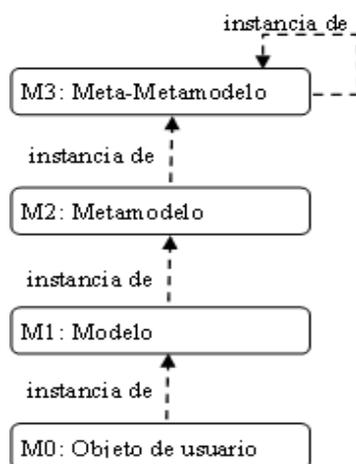


Figura 1. Arquitectura de Meta-modelado de MDA

En la *Figura 1. Arquitectura de Meta-modelado de MDA*, se puede ver esta arquitectura con los niveles de abstracción denominados M0, M1, M2 y M3. La estructura, términos, notaciones, sintaxis, semántica y reglas de integridad que son usados

para expresar un Modelo (M1), constituyen el lenguaje de modelado el cual, en su representación como modelo, es llamado Metamodelo (M2). Por esta razón un Modelo (M1) es considerado una instancia de un Metamodelo (M2). Dado que un Metamodelo (M2) es también un modelo, requiere también un lenguaje, un “Lenguaje de Meta-modelado” el cual, en su representación como modelo, recibe la denominación de Meta-Metamodelo (M3). Por esta razón un Metamodelo (M2) es considerado una instancia de un Meta-Metamodelo (M3). En el otro extremo, un Modelo (M1) también expresa un lenguaje de modelado, en este caso utilizado para expresar abstracciones de una realidad denominadas Objeto de Usuario (M0), razón por la cual una abstracción de una realidad, un Objeto de Usuario (M0), es considerada una instancia de un modelo (M1).

El lenguaje, en el nivel M3 (Meta-Metamodelo) utilizado para definir metamodelos, como por ejemplo el metamodelo UML, es MetaObject Facility (MOF) [8], cuya representación como modelo es un Meta-Metamodelo. A pesar de que es posible extender esta arquitectura con niveles más abstractos aún, para OMG no hay nuevo meta-nivel encima de MOF. Es decir que, en este punto, OMG decidió que el lenguaje utilizado para definir construcciones que especifiquen las reglas para la creación de ocurrencias en MOF, sea el propio MOF, el cual está definido como una extensión del metamodelo UML Core [8].

Arquitectura del Proceso de Desarrollo

Modelar el proceso de desarrollo, claramente constituye uno de los principales requisitos del proceso de maduración de la Ingeniería de Software. Los distintos modelos de ciclo de vida de desarrollo definidos a lo largo del proceso evolutivo han coincidido en algunas o todas y en el significado de un conjunto de categorías de actividades que han incorporado de diferentes maneras a sus respectivos modelos. Estas categorías de actividades son: modelado del problema, análisis del problema, diseño de la solución y construcción de la solución. La estrategia del enfoque MDA se apoya en considerar a los artefactos, resultantes de la ejecución de esas actividades del proceso de desarrollo, como modelos y en ordenar estos artefactos (modelos) en una arquitectura (dirigida por modelos) basada en niveles de abstracción.

En primer lugar, esta estrategia categoriza los modelos según su orientación en [7]:

- **Modelos de negocio o de dominio:** tienen por objetivo construir modelos del dominio asociados a la recolección de requisitos y son

históricamente denominados en OMG como Computational Independent Model (CIM).

- **Modelos de sistemas lógicos:** tienen por objetivo construir modelos lógicos orientados tanto al modelado del problema como del diseño de la solución y son categorizados como:
 - Modelos Independientes de la plataforma (PIM): orientados al modelado lógico del problema sin incluir decisiones vinculadas a la plataforma.
 - Modelos Dependientes de la plataforma (PSM): orientados al modelado del diseño de la solución en términos de la plataforma en la que se realizará el sistema.
- **Modelos de implementación:** tienen por objetivo modelar la forma en que un sistema o subsistema particular es implementado. Por estar ligados a una plataforma particular, frecuentemente son considerados como PSM. Sin embargo, algunos los consideran en una categoría específica, Modelos Específicos de Implementación (ISM).

Transformaciones de modelos en MDA

El enfoque MDA completa el modelo del proceso de desarrollo integrando los artefactos (modelos) y las actividades de desarrollo que los producen mediante el recurso de transformación. Estas transformaciones son definidas entre metamodelos para ser aplicadas a la transformación de uno o más modelos de entrada en otro de salida sin importar los niveles de abstracción definidos para la arquitectura del proceso de desarrollo. Dado que ambas arquitecturas son ortogonales, una transformación puede ser utilizada para producir una representación a partir de otra, o para cruzar capas arquitecturales.

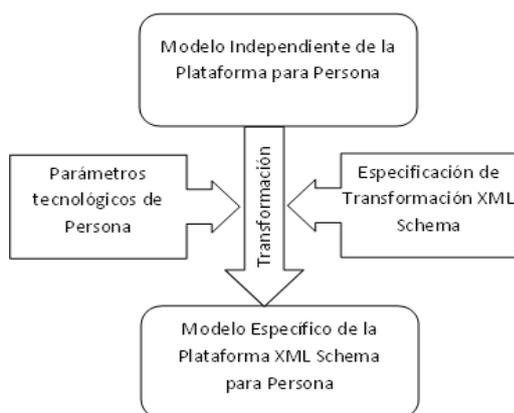


Figura 2. Ejemplo de Transformación Parametrizada [7]

Asumiendo las definiciones PIM y PSM introducidas en *Arquitectura del Proceso de Desarrollo*, la Figura 2. *Ejemplo de Transformación Parametrizada* presenta un ejemplo de un patrón de transformación, tomado de la guía “MDA Guide Version 2.0” [7], ilustra este concepto a través de los siguientes componentes:

- **Modelo independiente de la plataforma para Persona (PIM):** el modelo de información lógica representando información de persona.
- **Especificación de transformación XML Schema:** un componente reusable ejecutado por la transformación. Este componente es específico de XML Schema pero permite algunas opciones.
- **Parámetros tecnológicos de Persona:** una pequeña especificación que selecciona XML Schema como la tecnología deseada y proporciona cualquier parámetro requerido por el componente de transformación.
- **Transformación:** la transformación real que produce el PSM XML Schema para Persona.
- **Modelo específico de la plataforma XML Schema para Persona:** una representación XML Schema para Persona como está definida por el PIM.

Diseño de Software

Desde 2012 hasta la fecha, en el marco del proyecto, se desarrollan diferentes actividades de investigación orientadas a proveer más y mejores soportes al proceso de Diseño de Software. En particular, actividades desarrolladas en el marco de la línea de investigación presentada en la WICC 2015 [1], arrojaron como resultado una definición de Diseño de Software que heredó representatividad, estructura y organización, de la propuesta de definición del concepto de diseño formulada por Wand y Ralph [9]; y heredó especificidad en el dominio de Ingeniería de Software, de la propuesta de definición formulada por Buschmann et al. [10]. Además, extiende el concepto tradicional de Diseño de Software agregando más estructura y organización, incorporando conocimiento del enfoque de Patrones de Diseño de Software en la Arquitectura de Software resultante y proyectando luego esa estructura y organización en el proceso de Diseño de Software.

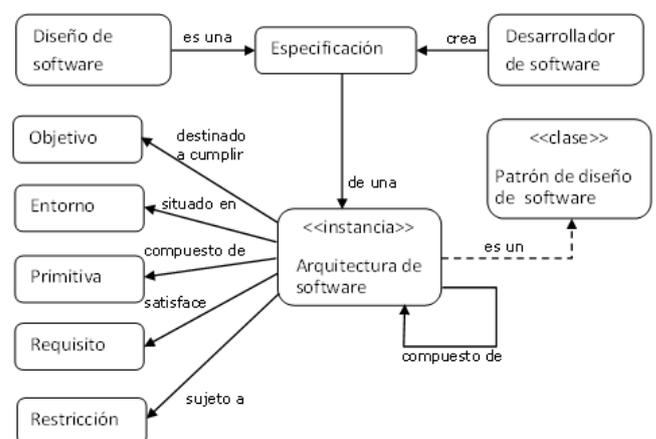


Figura 3. Modelo conceptual de diseño de software extendido con patrones [1] [9]

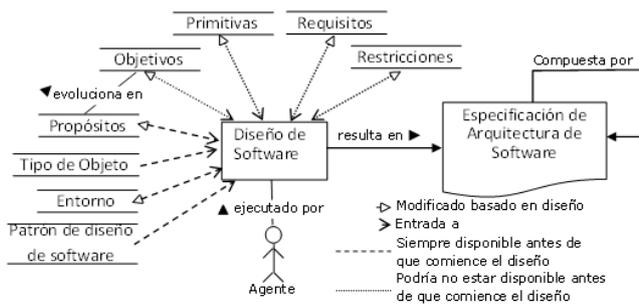


Figura 4. Nivel contextual del modelo conceptual de diseño extendido con patrones [1] [9]

Las dos figuras precedentes, presentan la definición utilizando un modelo conceptual y uno contextual de esta definición de Diseño de Software.

Como se observa en la *Figura 3. Modelo conceptual de diseño de software extendido con patrones*, la definición reconoce al objeto de diseño como Arquitectura de Software, pero además:

- La definición idealmente supone que cada parte que integra la Arquitectura de Software de una aplicación, puede ser obtenida como una instancia de un Patrón de Diseño de Software. Si bien no es ésta la realidad del estado actual, ni hay certeza de que lo vaya a ser en un estado futuro, sí hay importante evidencia de su efectividad en un buen número de casos. Por lo tanto, asumir sólo este tipo de casos, simplifica el análisis y no perjudica las conclusiones del trabajo, por el contrario, favorece su desarrollo despejando estos casos de interés y permitiendo la concentración en ellos sin comprometer la validez de los resultados. La pretensión aquí precisamente, consiste en distinguir el fuerte grado de dependencia (ser instancia de) del proceso de Diseño de Software respecto de Patrones de Diseño de Software, en comparación con el resto de las dependencias.
- La estructura que la definición reconoce en el objeto de diseño como Arquitectura de Software, es reforzada con la asociación 'compuesto de', la cual indica la capacidad que las Arquitecturas de Software tienen para componerse entre sí en la arquitectura del sistema. Un aspecto notable es la coincidencia de esta capacidad con la capacidad equivalente de composición entre patrones.

En la *Figura 4. Nivel contextual del modelo conceptual de diseño extendido con patrones*, este conocimiento de Patrones de Diseño de Software es inyectado como parámetro en el proceso influenciando la toma de decisiones requerida para producir la especificación de la Arquitectura de Software.

Línea de Investigación y Desarrollo

Como anticipamos en la *Introducción*, OMG a través del **Model Driven Architecture (MDA)**, aporta una meta-arquitectura para el proceso de desarrollo que cubre virtualmente la mayoría de las instancias de procesos de desarrollo posibles pero que adolece aún de un grado de abstracción que dificulta (o imposibilita) la derivación de soportes efectivos para actividades concretas de desarrollo de software como lo es la de síntesis y evaluación de alternativas para la toma de decisiones en el proceso de *Diseño de Software*.

Por este motivo presentamos aquí una línea de investigación la cual se propone especializar esta meta-arquitectura extendiéndola con resultados obtenidos en trabajos previos [1] [4] [5], de manera tal que el proceso de instanciación sea capaz de derivar arquitecturas más especializadas en el proceso de *Diseño de Software*.

La línea de investigación se fundamenta en la hipótesis de que si disponemos de una definición de *Diseño de Software* que incorpore la influencia de una representación estándar de conocimiento del dominio en la actividad de síntesis y evaluación de alternativas para la toma de decisiones y además, el tipo del resultado del proceso, define para éste una estructura compatible con esa definición, podemos extender la meta-arquitectura con este conocimiento, de manera tal que la instanciación de procesos de desarrollo de software sea capaz de derivar arquitecturas más especializadas, en particular en la actividad de síntesis y evaluación de alternativas para la toma de decisiones.

Presentamos a continuación, una versión extendida del recurso de *Transformaciones* aportado por el **Model Driven Architecture (MDA)** el cual expresa la hipótesis de trabajo de la línea de investigación propuesta.

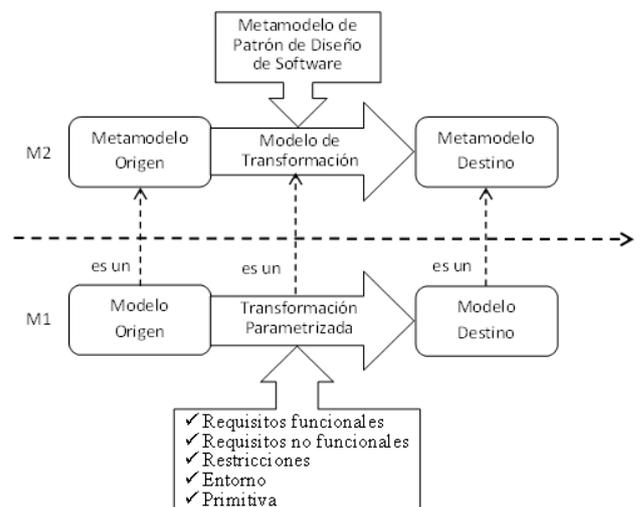


Figura 5. Modelo extendido de transformación

Resultados y Objetivos

En la *Figura 5. Modelo extendido de transformación*, se ve el proceso de transformación claramente influenciado por el conocimiento de la definición de *Diseño de Software* que hemos referido.

Son variados e interesantes los trabajos que se desarrollan en el marco de la línea de investigación y desarrollo presentada, muchos de los cuales se encuentran en curso en forma de trabajos de tesis de grado, trabajos de tesis de maestría, trabajos de publicación de resultados y trabajos de laboratorio en las carreras de Licenciatura en Ciencias de la Computación y de Ingeniería de Sistemas de la Universidad Nacional de San Luis. Entre estos trabajos podemos destacar:

- 1) Finalización de una tesis de la Maestría en Ingeniería de Software de la Universidad Nacional de San Luis, a ser presentada en los próximos meses, que presenta una propuesta integral de una meta-arquitectura que sintetiza la mayoría de los resultados hasta aquí obtenidos
- 2) En el presente año se inician trabajos de tesis en las carreras de Licenciatura en Ciencias de la Computación y de Ingeniería de Sistemas que se ocupan del desarrollo de mecanismos de instanciación Arquitecturas de Software a partir de Patrones de Diseño de Software basados en técnicas Orientadas a Objetos en un caso y en técnicas de XML Schemas en otro.
- 3) Este año se realizarán publicaciones de resultados de trabajos que se ocupan de la investigación de propiedades acerca de la relación de instancia entre Patrones de Diseño de Software y las Arquitecturas de Software, como resultado de efectuar extrapolaciones a dominios como Orientación a Objetos, XML Schema y Gramáticas de Lenguajes.
- 4) Actualmente se encuentra en curso en el marco de la línea de investigación y desarrollo trabajos que tienen por objetivo:
 - a. El desarrollo del metamodelo para Patrones de Diseño de Software.
 - b. El desarrollo un XML Schema para la representación XMI de instancias del metamodelo de Patrones de Diseño de Software
 - c. El diseño de un módulo reutilizable por las transformaciones para extenderlas con la capacidad de procesar representaciones del metamodelo de Patrones de Diseño de Software a la luz de información del dominio recibida como parámetro u obtenida por

medio de consultas a otros objetos del dominio

Para el presente año, el objetivo es recolectar e integrar resultados obtenidos por la finalización de algunos de los trabajos en curso o de resultados intermedios de los mismos y plasmarlos en publicaciones correspondientes al próximo año.

Formación de Recursos Humanos

En el área de la temática planteada, se desarrollan diferentes actividades que van desde la inclusión de trabajos prácticos, laboratorios y ejercicios de investigación, en materias de las carreras afines hasta el desarrollo de tesis y trabajos finales, algunos finalizados y otros en curso, en las mismas.

En particular, durante el año 2015, finalizó una tesis de maestría que experimentó con este enfoque y que será presentada en los próximos meses.

Referencias

- [1] L. Roqué Fourcade y L. Arakaki, Soporte a la actividad de diseño basado en patrones de diseño, XVII WICC, Salta, 2015.
- [2] Object Management Group, MDA - The Architecture of Choice for a Changing World, <http://www.omg.org/mda>.
- [3] Object Management Group, About OMG, <http://www.omg.org/gettingstarted/gettingstartindex.htm>.
- [4] L. Roqué Fourcade y L. Arakaki, Derivando el diseño a partir de especificaciones de requisitos basadas en Casos de Uso, XIV WICC, Posadas, Misiones, Argentina, 2012.
- [5] L. Roqué Fourcade, Derivando el Diseño a Partir de Especificaciones de Requisitos Basadas en Casos de Uso, 13th ASSE, La Plata, Buenos Aires, Argentina, 2012.
- [6] O. M. Group, Documents associated with Unified Modeling Language™ (UML®), v2.4.1, <http://schema.omg.org/spec/UML/2.4.1>.
- [7] Object Management Group, OMG Document -- ormsc/14-06-01 (MDA Guide revision 2.0), <http://www.omg.org/cgi-bin/doc?ormsc/14-06-01>.
- [8] Object Management Group, OMG's MetaObject Facility, <http://www.omg.org/mof/>.
- [9] P. Ralph y Y. Wand, de A Proposal for a Formal Definition of the Design Concept, Design Requirements Workshop, Berlin, 2009.
- [10] F. Buschmann, R. Meunier, H. Rohnert, P. Sommerlad y M. Stal, Pattern-Oriented Software Architecture: A System of Patterns, Chichester, England: Wiley, 2001.

Verificación en Alloy de modelos y metamodelos específicos del dominio

Ana Garis¹, Alejandro Sánchez¹

¹ Universidad Nacional de San Luis
Ejército de los Andes 950, San Luis, Argentina
{agaris, asanchez}@unsl.edu.ar

Resumen

La verificación de modelos de sistemas de software es una actividad clave para mejorar la calidad del producto final.

Numerosos lenguajes específicos del dominio han sido creados para favorecer la definición de modelos ajustados a un dominio o área particular. Sin embargo, la verificación de estos modelos, frecuentemente es dejada de lado como actividad complementaria al modelado.

Esta línea de investigación, desarrollo e innovación se orienta a generalizar un mecanismo basado en Alloy, para la especificación y verificación de modelos y metamodelos específicos del dominio. Alloy es un lenguaje formal, soportado por una amigable herramienta de verificación y validación. Las características de Alloy pueden ser aprovechadas para establecer un enfoque que permita garantizar la calidad de los modelos de sistemas de software específicos.

Palabras clave: Alloy, DSL, Verificación

Contexto

Esta línea de Investigación, desarrollo e innovación (I-D-I) se inserta en el proyecto “Ingeniería de Software” coordinado por la facultad de Ciencias Físico – Matemáticas y Naturales de la Universidad Nacional de San Luis (UNSL).

Introducción

Un *lenguaje específico de dominio* (DSL, por sus siglas en inglés) restringe sus primitivas a un problema específico con el objetivo de facilitar a los expertos el desarrollo de modelos [1]. Los DSLs tienen gran importancia dentro del enfoque *model driven architecture* (MDA) [2], ya que se utilizan para anotar modelos independientes de la plataforma, con el objetivo de transformarlos a uno o mas modelos específicos de la plataforma. De esta forma, siguiendo con la propuesta MDA, los modelos son sucesivamente refinados hasta obtener especificaciones cada vez mas concretas (código), alcanzando el producto final.

La verificación formal de los modelos en MDA permitiría incrementar la calidad del producto final. La verificación basada en métodos formales provee un poderoso mecanismo para detectar inconsistencias y chequear propiedades deseadas. Sin embargo, es infrecuente encontrar DSLs con herramientas que soporten la verificación formal de sus modelos, o que cuando lo hagan, esta verificación pueda ser complementada en sus aspectos menos fuertes.

Alloy [3] es un lenguaje formal, que incluye una amigable herramienta de *verificación y validación* (V&V) soportada por un analizador SAT. La especificación de una DSL en Alloy, permitirá aprovechar el potencial de Alloy para su V&V.

En trabajos anteriores, mostramos un enfoque basado en Alloy para la especificación, V&V de los aspectos estructurales que definen el metamodelo de un DSL [4,5]. Esto permite establecer una representación mas precisa del DSL propiamente dicho.

La línea de I-D-I propone un mecanismo general centrado en Alloy para la verificación de modelos y metamodelos específicos del dominio. Primero se valida y verifica el metamodelo de una DSL, y luego los modelos de esta. De esta forma, se extiende el enfoque en trabajos anteriores, ya que luego de realizar la V&V del metamodelo, nos centramos en desarrollar una guía para la verificación de sus modelos. El seguimiento de esta guía de habilitará la especificación de modelos más robustos, que llevará a la obtención de productos de software de alta calidad.

Líneas de Investigación, Desarrollo e Innovación

Los ejes de esta línea de I-D-I se detallan a continuación.

- La definición de una guía para la especificación y verificación con Alloy de modelos específicos del dominio. La guía debería establecer un proceso que indique cómo obtener una especificación Alloy que represente a la especificación en la DSL, y de qué manera éste debe ser verificado.

- La integración de métodos formales para la verificación de modelos específicos del dominio. Si bien, en su gran mayoría los DSLs carecen de una base formal que habilite su verificación, algunos de ellos presentan un sustento formal. En estos últimos casos, el enfoque propuesto debería integrar las técnicas y herramientas de verificación ya soportadas, con las que posee Alloy,

potenciando la verificación formal de modelos específicos de dominio. El Analizador de Alloy, permite entre otras cosas, buscar instancias de los modelos, y detectar contraejemplos, utilizando la técnica Bounded Model Checking [6].

- La definición de casos de estudio. Diferentes casos de estudio permitirán testear la propuesta planteada. Esto es, una guía para la especificación, V&V de modelos específicos del dominio usando Alloy.

Resultados y Objetivos

En anteriores trabajos, planteamos llevar a cabo la especificación, validación y verificación del metamodelo de DSLs. Como caso de estudio tomamos a Archery – un lenguaje diseñado para el modelado, análisis y verificación de arquitecturas de software [7].

El objetivo de la presente línea de I-D-I es definir un mecanismo que permita obtener modelos en Alloy a partir de especificaciones en DSLs, utilizando el metamodelo de dicha DSL en Alloy (ya obtenido previamente), y permitiendo la V&V del modelo con el analizador.

Para llevar a cabo esta tarea, utilizaremos nuevamente como caso de estudio a Archery. Dada las características formales de este lenguaje, la especificación en Alloy puede potenciar la V&V, enriqueciéndola con los atributos que provee el analizador.

La verificación en Archery esta basado en un model checking exhaustivo, que analiza todos los estados posibles del sistema para chequear propiedades deseadas cuando ésta no es satisfecha. Sin embargo, los contraejemplos brindados por el fundamento formal de Archery son de difícil interpretación. Por otro lado, tiene asociado limitaciones tales como la explosión de estados, condicionando el

modelado sólo a la representación de sistemas de estado finito.

Por otro lado, Alloy lleva a cabo un bounded model checking; es decir, establece una aproximación de estados alcanzables hasta un límite determinado de antemano. Esto alivia el problema de la explosión de estados, y permite descubrir errores rápidamente, aunque no sea posible probar la ausencia de errores para todos los casos. Por esta razón, Alloy necesita ser combinado con otros métodos formales en la verificación de sistemas críticos, en donde el error debe ser totalmente descartado.

En la literatura, pueden encontrarse trabajos donde se muestra la conexión entre Alloy y otros enfoques formales [8, 9, 10, 11].

Entre los objetivos de la presente línea de investigación se incluye elaborar un marco de trabajo para explotar la integración de Alloy con otros métodos formales, que a su vez, sean los fundamentos semánticos de DSLs. El framework debería indicar cómo utilizar Alloy combinado con el enfoque formal que soporte el DSL. Para llevar a cabo esta tarea, se deben caracterizar diferentes métodos formales con respecto a su integración con Alloy. De esta forma, dependiendo de la herramienta formal de la DSL, la guía permitirá integrar los enfoques.

Adicionalmente, se deben establecer otros casos de estudio que permitan fortalecer la propuesta aquí planteada. En este sentido, deben evaluarse diversas DSLs existentes en la actualidad.

Formación de Recursos Humanos

Dentro de la línea de investigación se desarrollan trabajos de tesis correspondiente al doctorado en Cs. de la Computación, Universidad Nacional de San Luis.

Referencias

- [1] Arie van Deursen, Paul Klint, and Joost Visser. “Domain-specific languages: An annotated bibliography”. SIGPLAN Not., 35(6):26–36, 2000.
- [2] OMG: MDA Guide, version 1.0.1, 2003.
- [2] OMG: Meta Object Facility Core, v2.4.2, 2014.
- [3] Daniel Jackson. Software Abstractions: Logic, Language, and Analysis. MIT Press, edición revisada, 2012.
- [4] Ana Garis, Alejandro Sanchez. “Verification and Validation of Domain Specification Languages using Alloy”. Proceedings of the XXI Congreso Argentino de Ciencias de la Computación (CACIC 2015), ISBN 978-987-3724-37-4, pp. 589-598, 2015.
- [5] Ana Garis, Alejandro Sánchez. "Especificación Formal de Lenguajes Específicos del Dominio utilizando Alloy", Proceedings of the XVII Workshop de Investigadores en Ciencias de la Computación (WICC2015), ISBN 978-987-633-134-0, id 6898, 2015.
- [6] Edmund Clarke, Armin Biere, Richard Raimi, y Yunshan Zhu, “Bounded Model Checking Using Satisfiability Solving”, Journal of Formal Methods in System Design, Kluwer Academic Publishers, pp. 7-34, 2001.
- [7] Alejandro Sanchez, Luis Barbosa y Daniel Riesco. “A language for behavioural modelling of architectural patterns”. Proceedings of the Third Workshop on Behavioural Modelling, pp. 17-24, 2011.
- [8] M. Ferreira, S. Silva, and J. Oliveira. “Verifying Intel flash file system core specification”. In Modelling and Analysis

in VDM: Proceedings of the Fourth VDM/Overture Workshop, 2008.

[9] Paulo Matos, João Marques-Silva. “Model checking Event-B by encoding into Alloy”. Proceedings of the First International Conference, ABZ. Volume 5238/2008, page 346. 2008.

[10] Renato Neves , Alexandre Madeira, Manuel Martins, Luís Barbosa. “An Institution for Alloy and Its Translation to Second-Order Logic”. In Integration of Reusable Systems: Advances in Intelligent Systems and Computing, Springer, pp 45-75. 2014.

[11] Miguel A Ferreira, Jose N. Oliveira, “Variations on an Alloy-centric tool-chain in verifying a journaled file system model”, Techn. Report DI-CCTC-10-07. 2010.

FORMULACIÓN DE UN MODELO DE PROCESO PARA INGENIERÍA DEL CONOCIMIENTO

Hernán Amatriain, Ezequiel Baldizzoni, Sebastian Martins, Santiago Bianco,
Eduardo Diez, Ramón García-Martínez

Laboratorio de Investigación y Desarrollo en Sistemas de Inteligencia Artificial
Grupo Investigación en Sistemas de Información

Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús. Argentina. Tel +54 11 5533 5600 Ext. 5194
hamatriain@gmail.com, rgm1960@yahoo.com

RESUMEN

Las metodologías de desarrollo de sistemas basados en conocimiento existentes provistas por la Ingeniería de Conocimiento, se centran en el proceso de desarrollo de Bases de Conocimiento. Sin embargo no existe una visión de proceso que identifique fases, tareas, técnicas de representación y procedimientos de ejecución de las tareas; ni que permitan soportar de forma completa e integrada las actividades de administración y desarrollo de Proyectos de Ingeniería de Conocimiento. En este contexto, esta investigación propone cubrir la vacancia desarrollando un Modelo de Proceso para Proyectos de Ingeniería del Conocimiento que integre en fases las actividades y técnicas desarrolladas para la administración y desarrollo de este tipo de proyectos.

Palabras clave: Modelo de Proceso, Ingeniería del Conocimiento, Sistemas Basados en Conocimiento

CONTEXTO

El proyecto: [a] inicia una línea de trabajo en el campo de la mejora de procesos productivos de Sistemas Basados en Conocimiento dentro del campo de la Informática; [b] articula la Línea de Investigación Prioritaria “3. Desarrollos Informáticos” del Instituto de Economía, Producción y Trabajo, aprobada por Resolución Consejo Superior UNLa N° 113/14, promoviendo la mejora de los sistemas productivos, en particular los utilizados en la producción de tecnologías especiales; y [c] responde a los lineamientos estratégicos de la CADENA DE VALOR DEL SOFTWARE Y SERVICIOS INFORMÁTICOS que establece el Plan Industrial 2020 del Ministerio Industria [MI, 2014], promoviendo la mejora del sistema productivo de la industria del Software con foco en sistemas software especiales.

INTRODUCCIÓN

Se pueden definir los Sistemas Basados en Conocimiento (SBC) como una clase de sistemas que son capaces de: aconsejar, categorizar, analizar, comunicar, consultar, diseñar, diagnosticar, explicar, explorar, formar conceptos, interpretar, justificar, planificar ; son en suma, programas capaces de manejar problemas que normalmente requieren para su resolución la intervención humana especializada [García-Martínez y Britos, 2004]. La Ingeniería del Conocimiento es la subdisciplina Informática que se ocupa de los procesos constructivos de los Sistemas Basados en Conocimiento [Hayes y Parzen, 1997; Kang y Bahieel, 1990].

Los SBC son desarrollados con la ayuda de Expertos de Campo, los cuales revelan información acerca de aquellos procesos mentales, que le permiten solucionar los distintos problemas. El otro profesional interviniente es el Ingeniero de Conocimiento [Giarratano y Riley, 2004], cuya función específica es la de dar forma simbólica y automáticamente manipulable al conocimiento proporcionado por el Experto de Campo.

Los expertos normalmente solucionan problemas mal definidos y desestructurados, los cuales generalmente involucran diagnóstico o planificación, para resolverlos generalmente usan heurísticas, es decir métodos que determinan que parte de su experiencia son aplicables, estas heurísticas deben ser descubiertas por el Ingeniero de Conocimiento y programadas en el SBC [Kendal y Creen, 2006].

Las siguientes características son deseables (aunque no siempre obtenibles) de los SBC [Menzies y Cukic, 1999; 2000]:

- [i] Aplican su experiencia de una manera eficiente para solucionar problemas, pudiendo realizar inferencias a partir de datos incompletos o inciertos.
- [ii] Explican y justifican lo que están haciendo.

- [iii] Se comunican con otros expertos y adquieren nuevos conocimientos.
- [iv] Reestructuran y reorganizan el conocimiento.
- [v] Pueden quebrantar reglas, es decir, interpretan simultáneamente el espíritu y la letra de de las mismas.
- [vi] Determinan cuando un problema está en el dominio de su experiencia, conocido como determinación de la relevancia del problema.

La Ingeniería del Conocimiento tiene una bien establecida tradición metodológica de mas de un cuarto de siglo [Grover,1983; Greenwell, 1988; Debenham, 1989; Brulé y Blount, 1989; Meyer y Booker, 1991; Garcia-Martinez, 1992; 1994]. La concepción ingenieril que ha tomado los procesos constructivos de las aplicación informáticas en los ultimos lustros sostiene la necesidad de una revisión de las practicas constructivas de sistemas basados en conocimiento.

Con base en que la Ingeniería de Software ha sido definida en el SWEBOOK [Abran et al., 2004] como: “la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento de software, y el estudio de estos enfoques, es decir, la aplicación de la ingeniería al software”; se conviene en proponer como definición de la Ingeniería del Conocimiento [Liebowitz, 2001; Konar y Jain, 2005; Hauge et al., 2006] como la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo de sistemas Basados en Conocimiento, y el estudio de este enfoque, es decir, la aplicación de la ingeniería a la construcción de SBC.

En este contexto, emerge la necesidad de reformular los procesos constructivos de SBC a partir de una revisión de las practicas constructivas de los SBC a la luz de las nuevas concepciones sobre procesos [Hossian, 2012].

PREGUNTAS PROBLEMA, HIPÓTESIS Y OBJETIVOS GENERALES Y ESPECÍFICOS

Preguntas Problema:

¿Se puede cubrir la vacancia de Modelo de Proceso para Proyectos de Ingeniería de Conocimiento que integre en fases las actividades y técnicas desarrolladas para el control, gestión y desarrollo de proyectos de Sistemas Basados en Conocimiento?

Hipótesis:

Hipótesis I:

Las metodologías de desarrollo de proyectos de Ingeniería de Conocimiento existentes, se centran en aspectos de desarrollo vinculados a la construcción y validación de la Base de Conocimiento del

Sistema. Sin embargo no existe una visión de proceso que identifique fases, tareas, técnicas de representación y procedimientos de ejecución de la tarea; que permita sistematizar la concreción del proyecto.

Hipótesis II:

Las metodologías de desarrollo de proyectos de Ingeniería de Conocimiento existentes, se centran en el proceso de desarrollo de Sistemas Basados en Conocimiento. Sin embargo, las metodologías usuales para este tipo de proyectos no contemplan el nivel de administración del control y la gestión de proyecto que los nuevos modelos de proceso de software prevén.

Objetivo General:

El objetivo de este proyecto es sistematizar el conocimiento existente sobre Ingeniería de Conocimiento, y formular un Modelo de Proceso de Gestión y Desarrollo de Proyectos de Ingeniería de Conocimiento: que: [a] integre las actividades en fases: y [b] para cada actividad identifique o desarrolle los métodos, las técnicas y las herramientas asociadas a la ejecución de aquella.

Objetivos Específicos:

Objetivo específico vinculado a la Hipótesis I (OE1):

Desarrollar un modelo de proceso de desarrollo de Proyectos de Ingeniería de Conocimiento que identifique fases, tareas, técnicas de representación y procedimientos de ejecución de la tarea que permita sistematizar el desarrollo de proyectos en el área.

Objetivo específico vinculado a la Hipótesis II (OE2):

Desarrollar un modelo de proceso de control y gestión de Proyectos de Ingeniería de Conocimiento que que identifique fases, tareas, técnicas de representación y procedimientos de ejecución de la tarea; que permita sistematizar el control y la gestión de proyectos en el área.

Objetivo específico vinculado a las Hipótesis I y II (OE3):

Desarrollar un modelo de Proceso Integrado de Proyectos de Ingeniería de Conocimiento que articule los modelos de Proceso de Control y Gestión y de Desarrollo para este tipo de proyectos.

METODOLOGÍA DE TRABAJO

Para construir el conocimiento asociado al presente proyecto de investigación, se seguirá un enfoque de investigación clásico [Riveros y Rosas, 1985; Creswell, 2002] con énfasis en la producción de

tecnologías [Sábato y Mackenzie, 1982]; identificando métodos, materiales y abordaje metodológico necesarios para desarrollar el proyecto:

Métodos:

Revisiones Sistemáticas:

Las revisiones sistemáticas [Argimón, 2004] de artículos científicos siguen un método explícito para resumir la información sobre determinado tema o problema. Se diferencia de las revisiones narrativas en que provienen de una pregunta estructurada y de un protocolo previamente realizado.

Prototipado Evolutivo Experimental (Método de la Ingeniería):

El prototipado evolutivo experimental [Basili, 1993] consiste en desarrollar una solución inicial para un determinado problema, generando su refinamiento de manera evolutiva por prueba de aplicación de dicha solución a casos de estudio (problemáticas) de complejidad creciente. El proceso de refinamiento concluye al estabilizarse el prototipo en evolución.

Materiales:

Para el desarrollo de los formalismos y procesos propuestos se utilizarán:

- Formalismos de modelado conceptual usuales en la Ingeniería de Software [Rumbaugh et al., 1999; Jacobson et al., 2013] y en la Ingeniería del Conocimiento [García-Martínez y Britos, 2004].
- Modelos de Proceso usuales en Ingeniería de Software [IEEE, 1997; ANSI/IEEE, 2007; Oktaba et al., 2007].

Abordaje Metodológico:

Para el desarrollo de esta propuesta técnica se han previsto utilizar las siguientes metodologías de investigación y desarrollo:

Para el Objetivo OE1 se propone: (i) realizar una investigación documental exploratoria sobre modelo de proceso de desarrollo de proyectos existentes; (ii) identificar casos de estudio y casos de validación en el área de desarrollo de Proyectos de Ingeniería de Conocimiento; (iii) desarrollar mediante la metodología de prototipado evolutivo un modelo de proceso de desarrollo de Proyectos de Ingeniería de Conocimiento que identifique: fases, tareas, técnicas de representación y procedimientos de ejecución de cada tarea; y (iv) realizar pruebas de concepto en los casos de estudio y casos de validación identificados que corroboren el modelo de proceso propuesto.

Para el Objetivo OE2 se propone: (i) realizar una investigación documental exploratoria sobre modelo de proceso de administración del control y la gestión de proyectos existentes; (ii) identificar casos de

estudio y casos de validación en el área control y gestión de Proyectos de Ingeniería de Conocimiento; (iii) desarrollar mediante la metodología de prototipado evolutivo un modelo de proceso de control y gestión de Proyectos de Ingeniería de Conocimiento que identifique: fases, tareas, técnicas de representación y procedimientos de ejecución de cada tarea; y (iv) realizar pruebas de concepto en los casos de estudio y casos de validación identificados que corroboren el modelo de proceso propuesto.

Para el Objetivo OE3 se propone: (i) identificar precedencia de actividades entre el proceso de control y gestión y el proceso de desarrollo; (ii) identificar las articulaciones entre ambos modelos de proceso derivadas de las precedencias de las actividades identificadas; (iii) con base en las articulaciones identificadas, desarrollar mediante la metodología de prototipado evolutivo un modelo de proceso integrado de proyectos de ingeniería de explotación de información; y (iv) realizar pruebas de concepto en los casos de estudio y casos de validación identificados que corroboren el modelo de proceso propuesto.

RESULTADOS OBTENIDOS/ESPERADOS

El proyecto prevé formular aportaciones conceptuales en el área de modelos de proceso para Ingeniería del Conocimiento. Contar con herramientas de este tipo permite explorar formas de desarrollo sistemático de procesos constructivos de sistemas expertos.

FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por dos investigadores formados, tres investigadores en formación, y un asesor en metodología de la investigación. En su marco se desarrollan una Tesis de Doctorado en Ciencias Informáticas y dos Tesis de Maestría en Tecnología Informática.

FINANCIAMIENTO

Las investigaciones que se proponen en esta comunicación cuentan con financiamiento como Proyecto 80020150200073LA de la Secretaría de Ciencia y Técnica de la Universidad Nacional de Lanús (Argentina).

REFERENCIAS

- ANSI/IEEE, 2007. Draft IEEE Standard for software and system test documentation. ANSI/IEEE Std P829-2007.
- Argimón, J. 2004. Métodos de Investigación Clínica y Epidemiológica. Elsevier. 84-8174-709-2.

- Basili 1993. The Experimental Paradigm in Software Engineering. En *Experimental Software Engineering Issues: Critical Assessment and Future Directions* (Ed. Rombach, H., Basili, V., Selby, R.). Lecture Notes in Computer Science, Vol. 706. ISBN 978-3-540-57092-9.
- Brulé, J. y Bount, A. 1989. *Knowledge Acquisition*. McGraw-Hill. New York.
- Creswell, J. 2002. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Prentice Hall. ISBN 10: 01-3613-550-1.
- Debenham, J. 1989. *Knowledge System Design*. Prentice Hall. Sidney.
- García Martínez, R. 1992. *Construcción de Sistemas Expertos*. 154 páginas. Imprenta del CEI-UBA. Argentina.
- García Martínez, R. 1994. *Adquisición de Conocimiento*. En Abecasis, S. y Heras, C. *Metodología de la Investigación*. 157 páginas. Editorial Nueva Librería. ISBN 950-9088-65-x.
- García Martínez, R. y Britos, P. 2004. *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería. ISBN 987-1104-15-4.
- Giarratano, J. y Riley, G. 2004. *Expert Systems: Principles and Programming*. PWS Publishing Company.
- Gómez, A., Juristo, N., Montes, C. Pazos, J. 1997. *Ingeniería de Conocimiento*. Editorial Centro de Estudio Ramón Areces. ISBN 84-8004-269-9.
- Greenwell, M. 1988. *Knowledge Engineering for Expert Systems*. Ellis Horwood Limited. Chichester.
- Grover, M. 1983. *A Pragmatic Knowledge Aquisition Methodology*. Proceedings VIII IJCAI. Estados Unidos.
- Hauge, O., Britos, P., García-Martínez, R. 2006. *Conceptualization Maturity Metrics for Expert Systems*. IFIP International Federation for Information Processing, Volume 217, *Artificial Intelligence in Theory and Practice*, ed. M. Bramer, (Boston: Springer), pp. 435-444.
- Hayes C. y Parzen. M. 1997. *QUEM: An achievement Test for Knowledge-Based Systems*, IEEE Transactions on Knowledge and Data Engineering, Vol. 9, No. 6, November/December 1997.
- Hossian, A. 2012. *Modelo de Proceso de Conceptualización de Requisitos*. Tesis Doctoral en Ciencias informáticas. Facultad de Informática. Universidad Nacional de La Plata.
- IEEE, 1997. *IEEE Standard for Developing Software Life Cycle Processes*. IEEE Std 1074-1997 (Revision of IEEE Std 1074-1995; Replaces IEEE Std 1074.1-1995)
- Kang, Y. & Bahieel, T. 1990. *A Tool for Detecting Expert Systems Errors*. AI Expert, 5(2): 42-51.
- Kendal, S. y Creen, M. 2006. *An Introduction to Knowledge Engineering*. Springer.
- Konar, A. y Jain, L. 2005. *Cognitive Engineering: A Distributed Approach to Machine Intelligence*. Springer.
- Liebowitz, J. 2001. *Knowledge Management: Learning from Knowledge Engineering*. CRC Press.
- Menzies, T. y Cukic, B. 1999. *On the Sufficiency of Limited Testing for Knowledge Based Systems*. Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence, Pages 431-440.
- Menzies, T. y Cukic, B. 2000. *Adequacy of Limited Testing for Knowledge Based Systems*. International Journal on Artificial Intelligence Tools 9(1): 153-172.
- Meyer, M. y Booker, J. 1991. *Eliciting and Analyzing Expert Judgement. A Practical Guide*. Academic Press. Londres.
- Oktaba, H., Garcia, F., Piattini, M., Ruiz, F., Pino, F., Alquicira, C. 2007. *Software Process Improvement: The Competisoft Project*. IEEE Computer, 40(10): 21-28. ISSN 0018-9162.
- Riveros, H. y Rosas, L. 1985. *El Método Científico Aplicado a las Ciencias Experimentales*. Editorial Trillas. México. ISBN 96-8243-893-4.
- Rumbaugh, J., Jacobson, I., Booch, G. 1999. *The Unified Modeling Language, Reference Manual*. Addison Wesley, ISBN-10: 02-0130-998-X.
- Sábato, J. y Mackenzie, M. 1982. *La Producción de Tecnología*. Editorial Nueva Imagen. México. ISBN 968-429-348-8.

Asistencia a la identificación de especies botánicas del NE del Chubut a través de aplicaciones basadas en ontologías

Renato Mazzanti^{1,2}, Gustavo Daniel Samec^{1,2}, Alejandro Bisigato³, Alejandro Sanchez^{1,4}

¹LINVI, Dpto. de Informática, Fac. de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco
+54-0280-4472885-int. 117 - Puerto Madryn, Argentina

²Unidad de Gestión de la Información – UGI CENPAT-CONICET
+54-0280-4450401-int. 1260 - Puerto Madryn, Argentina

³Instituto Patagónico para el Estudio de los Ecosistemas Continentales – IPEEC-CONICET
+54-0280-4450401-int. 1321 - Puerto Madryn, Argentina

⁴Departamento de Informática, Universidad Nacional de San Luis
+54 266 4520300 – San Luis, Argentina

e-mail: {renato, gsamec, bisigato}@cenpat-conicet.gob.ar, aljsanchez@gmail.com

Resumen

La determinación de especies botánicas dista de ser trivial. El enfoque mayormente aplicado consiste en utilizar claves dicotómicas que a través de una serie preestablecida de observaciones, ayudan al experto a determinar la identidad de una planta. Estas claves se materializan en voluminosas publicaciones que incluyen un extenso conjunto de especies, las que raramente coexisten en un sitio, y requieren reconocer caracteres sutiles, o el uso de instrumentos de laboratorio. El proceso es complejo, y resulta inviable cuando se debe completar en campo.

Esta línea de investigación, desarrollo e innovación apunta a definir y desarrollar aplicaciones basadas en ontologías para asistir en la determinación de especies a investigadores botánicos del NE de la provincia del Chubut. Estas aplicaciones deberán funcionar, tanto en laboratorio como en campo, sin seguir una secuencia preestablecida, ajustándose a las características reconocidas por el usuario.

Palabras clave: Bioinformática, Ontología, Morfología de Plantas, Botánica

Contexto

Esta línea de investigación se encuentra en su etapa inicial y será llevada adelante en el contexto del Laboratorio en Investigación en Informática (LINVI) del Departamento de Informática de la Universidad Patagónica San Juan Bosco (UNPSJB), en su sede de Puerto Madryn. Se relaciona con las líneas de I-D-I del LINVI, BigData y Ontologías, e involucra al Profesor de Ecología de Comunidades de la UNPSJB el Dr. Alejandro Bisigato.

Introducción

La identificación de especies de plantas suele ser un desafío incluso para los expertos. La situación es aún más desafiante cuando el proceso debe completarse en campo. Aunque existen

numerosas claves dicotómicas, éstas suelen abarcar grandes regiones e incluir cientos o miles de especies. Ello hace que las claves hagan referencia a sutiles caracteres para diferenciar especies emparentadas, las que pueden estar especificadas utilizando conceptos (o lenguaje) que resulte ajeno a un botánico menos experto por su especificidad o taxonomía subyacente, o que difícilmente puedan observarse en campo sin la ayuda de material óptico. Un ejemplo de estas claves es la Flora Patagónica (Correa 1969-1999), la que incluye varios tomos y nunca terminó de publicarse de modo impreso. Ello ha originado la publicación de numerosas guías locales, las que frecuentemente incluyen sólo algunos grupos de plantas y muchas veces, incluso carecen de claves para la determinación (Bertiller y Beeskow S/F, Green y Ferreyra 2011, Kröpfl y Vilasuso 2012, Quintana 2014).

En los últimos años se han desarrollado varias iniciativas en el dominio de la biología donde se utilizan distintos enfoques para acercar comunidades que realizan investigaciones en diferentes disciplinas tales como la genética, biología molecular, biología celular, taxonomía, botánica y genómica. Estos enfoques incluyen, por ejemplo, la utilización de vocabulario controlado, como en Plant Ontology Database¹, bases de datos online, como el caso de Flora Argentina², generada y mantenida por la autoridad taxonómica de la botánica para Argentina, o lenguajes y herramientas específicos, como es el caso de DELTA³ (DEscription Language for TAXonomy)

¹ <http://www.plantontology.org/>

² <http://www.floraargentina.edu.ar/>

³ <http://delta-intkey.com/>

(Dallwitz 1980), cuyas herramientas interactivas permiten seguir, a diferencia de las claves dicotómicas, la trayectoria que mejor se adapte a la información que dispone el usuario.

El uso de ontologías para la representación formal y compartida del conocimiento (Borst 1997, Gruber 1993) emerge también como una opción para atender a problemas como el enunciado (Fang y colegas 2004). Los lenguajes y tecnologías que soportan este enfoque poseen fundamentos sólidos en las lógicas descriptivas (Baader y colegas 2003). Por ejemplo, un razonador responderá la especie/subespecie de una instancia una vez se describan todas las características necesarias para tal clasificación.

En general, la conceptualización se formaliza en una red de ontologías relacionadas de distintos tipos. Una posible clasificación de las ontologías distingue entre ontologías de alto nivel, de tarea, de dominio, y de aplicación (Guarino 1998). Las ontologías de un dominio son esenciales en estos enfoques. Las entidades en estas se relacionarán con conceptos definidos en ontologías de alto nivel, como ser, por ejemplo, la ontología de unidades estándar de medida.

Por otro lado, las ontologías de aplicación refinan las de dominio para responder a preguntas/problemas específicos dentro del dominio, dando soporte al funcionamiento de una aplicación. Las relaciones en esta red de ontologías se establecen con tecnologías de correspondencia entre ontologías (Euzenat 2013).

Las ontologías están lejos de ser desarrolladas de manera anárquica. Existe un conjunto de metodologías para esto, incluyendo (Grüniger y Fox 1995, Uschold y King 1996, Bernaras y colegas 1996, Gómez-Pérez y colegas 1996, Staab y colegas 2001, Corcho y colegas 2003, Tautz y Wangenheim 1998). Muchas de

estas metodologías consideran la participación de expertos de dominio (no familiarizados con lenguajes de especificación de ontologías) en el proceso de desarrollo.

Líneas de Investigación, Desarrollo e Innovación

Esta línea de I-D-I procura definir y desarrollar prototipos de aplicaciones basadas en ontologías para asistir en la determinación de especie en el NE de la provincia de Chubut. El enfoque se inspira en la necesidad de brindar funcionalidad complementando la provista por DELTA, pero explotando la expresividad y herramientas disponibles para ontologías especificadas en lógicas descriptivas (o también llamadas de la web semántica). Los principales ejes de la línea son:

- Participación de expertos del dominio de la región como usuarios y fuente de conocimiento.
- Selección y ajuste de metodología de desarrollo de ontologías, y software asociado.
- Utilización herramientas para el desarrollo de ontologías especificadas en lenguajes basados en lógicas descriptivas.
-

Objetivos y resultados esperados

Con el objetivo general de asistir en la determinación de especie en el NE de la provincia de Chubut, a través de aplicaciones basadas en ontologías, se desarrollarán una serie de actividades que se esperan que permitan:

- Obtener una aproximación explícita e informal a la conceptualización para

responder a las preguntas del experto, a partir de las fuentes de conocimiento disponibles que describen la flora de la provincia, y descripciones de los expertos de la región.

- Desarrollar una red de ontologías a partir de la (re)utilización de ontologías existentes.
- Desarrollar aplicaciones basadas en ontologías que permitan al investigador identificar especies botánicas de la provincia en diversos escenarios de trabajo.

Formación de Recursos Humanos

Está previsto en esta línea de I-D-I continuar con la formación de integrantes del LINVI, incluyendo el desarrollo de dos tesis de Maestría en Ciencias de la Computación de la Universidad Nacional del Sur, y dos tesinas de grado de la Licenciatura en Informática de la UNPSJB.

Referencias

Atkinson B., Cooper L, Moore L, Preece J, Lingutla NTV, Todorovic S, Walls RL, Stockey R, Rothwell G, Smith B, Gandolfo MA, Stevenson DW, Jaiswal P. (2013) Plant Ontology, a controlled and structured plant vocabulary for all botanical disciplines. Botany Conference. New Orleans, LA

BERTILLER, MÓNICA y BEESKOW, ANA MARÍA. S/F. Las flores de la Patagonia. Alrededores de Puerto Madryn y Península Valdés. I. Arbustos y

Subarbustos. Área Científica de Zonas Áridas. Centro Nacional

Patagónico. CONICET. Puerto Madryn. Chubut. Argentina

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point?. *Data & knowledge engineering*, 46(1), 41-64.

Correa, M. N., 1969. Flora Patagónica. Parte II. Typhaceae a Orchidaceae. Colección Científica del INTA. Buenos Aires. 219 pp.

Correa, M. N., 1971. Flora Patagónica. Parte VII. Compositae. Colección Científica del INTA. Buenos Aires. 451 pp.

Correa, M. N., 1984. Flora Patagónica. Parte IVa. Dicotiledóneas Dialipétaleas (Salicaceae a Cruciferae). Colección Científica del INTA. Buenos Aires. 559 pp.

Correa, M. N., 1988. Flora Patagónica. Parte IVb. Dicotiledóneas Dialipétaleas (Droseraceae a Leguminosae). Colección Científica del INTA. Buenos Aires. 309 pp

Correa, M. N., 1988. Flora Patagónica. Parte V. Dicotiledóneas Dialipétaleas (Oxalidaceae a Cornaceae). Colección Científica del INTA. Buenos Aires. 381 pp.

Correa, M. N., 1999. Flora Patagónica. Parte VI. Dicotyledones. Gamopétaleas (Ericaceae a Calyceraceae). Colección Científica del INTA. Buenos Aires. 536 pp.

Dallwitz, M.J.. A general system for coding taxonomic descriptions. 1980.

Fang Gu, Cun-Gen Cao, Yue-Fei Sui, and Wen Tian. 2004. Domain-specific ontology of botany. *J. Comput. Sci. Technol.* 19, 2 (March 2004), 238-248.

DOI=<http://dx.doi.org/10.1007/BF02944802>

Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider (Eds.). 2003. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, New York, NY, USA. Borst, Willem Nico (1997). "Construction of Engineering Ontologies for Knowledge Sharing and Reuse". Tesis doct. Enschede.

Gómez-Pérez, A., Fernández, M., & Vicente, A. D. (1996). Towards a method to conceptualize domain ontologies.

Green L y Ferreyra M. 2011. Flores de la estepa patagónica. Vázquez Mazzini Editores, Buenos Aires.

Gruber, Thomas R. (1993). "A Translation Approach to Portable Ontology Specifications". En: *Knowl. Acquis.* 5.2, págs. 199-220. ISSN : 1042-8143.

Grüninger, M., & Fox, M. S. (1995). *Methodology for the Design and Evaluation of Ontologies*.

Guarino, N. (1998). *Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy*. 1st. Amsterdam, The Netherlands, The Netherlands: IOS Press. ISBN : 9051993994.

Jérôme Euzenat and Pavel Shvaiko. 2013. *Ontology Matching (2nd ed.)*. Springer Publishing Company, Incorporated

Kröpfl Ai & Villasuso Nm. 2012 *Guía para el Reconocimiento de Especies de los Pastizales del Monte Oriental de Patagonia*. San Carlos de Bariloche, INTA.

L. Watson and M. J. Dallwitz (1998 onwards). *DELTA Sample Data: Descriptions, Illustrations, Identification*,

and Information Retrieval. Version: 21st September 2000.

Quintana, DR 2014 Plantas de la Patagonia Árida. Nativas y exóticas del noroeste del Chubut. Fondo editorial de la Secretaría de Cultura de la Prov. del Chubut

Schreiber, A. T., Wielinga, B. J., Akkermans, J. M., & Van de Velde, W. (1994). CommonKADS: A comprehensive methodology for KBS development. Deliverable DM1 (Vol. 1). 2a KADSII.

Staab, S., Studer, R., Schnurr, H. P., & Sure, Y. (2001). Knowledge processes and ontologies. IEEE Intelligent systems, (1), 26-34.

Šimún, M., Andrejko, A., & Bieliková, M. (2007, September). Ontology-based models for personalized e-learning environment. In Proc. of 5th Int. Conf. on Emerging e-learning Technologies and Applications (ICETA'07).

Tautz, C., Wangenheim, G., & Refseno, C. (1998). A representation formalism for software engineering ontologies. Fraunhofer IESE-Report No, 15, 1-151.

Uschold, M., & King, M. (1995). Towards a methodology for building ontologies (pp. 15-30). Edinburgh: Artificial Intelligence Applications Institute, University of Edinburgh.

Simulando Proyectos de Desarrollo de Software que incorporan Test Driven Development

Diego Alberto Godoy^a, Henry Kotynski^b, Edgardo A. Belloni^c, Eduardo O.Sosa^d,
Mariela Gamonal^e, Nahuel Regueira^f

Centro de Investigación en Tecnologías de la Información y Comunicaciones (C.I.T.I.C.)
Departamento de Ingeniería y Ciencias de la Producción-Universidad Gastón Dachary
Av. López y Planes 6519- Posadas, Misiones, Argentina. Teléfono: +54-376-4438677

^a diegodoy@citic.ugd.edu.ar, ^b henrykotynski@citic.ugd.edu.ar, ^c ebelloni@ugd.edu.ar,
^d eduardo.sosa@citic.edu.ar, ^e mariela.gamonal@citic.edu.ar,
^f nahuel.regueira@citic.edu.ar

Resumen

En este artículo se presenta una línea de investigación denominada Simulación como Herramienta para la Mejora de los Procesos de Software Desarrollados con Metodologías Ágiles Utilizando Dinámica de Sistemas, cuyo objetivo es construir modelos de simulación utilizando la disciplina de dinámica de sistemas, que permitan estudiar los procesos de desarrollo de software ágiles gestionados por diversas metodologías y prácticas ágiles. Particularmente se presentan los avances realizados hasta ahora en relación a construcción de un Modelo de simulación para estudiar el comportamiento de proyectos de software que utilizan Test Driven Development (TDD) en su proceso de desarrollo. Para ello se ha realizado una encuesta en la comunidad de desarrolladores sobre el uso que le dan a TDD en sus proyectos y cuáles son las características de los mismos. Se presenta además el Subsistema de planificación de pruebas de aceptación, uno de los subsistemas más importantes dentro de modelo que se pretende construir.

Palabras claves: Administración de Proyectos de Desarrollo de Software; Test Driven Development Dinámica de Sistemas.

Contexto

El trabajo presentado en este artículo tiene como contexto marco el proyecto de investigación denominado “Simulación como Herramienta para la Mejora de los Procesos de Software Desarrollados con Metodologías Ágiles Utilizando Dinámica de Sistemas”, registrado actualmente en la Secretaría de Investigación y Desarrollo de la Universidad Gastón Dachary (UGD) [†] y radicado en el Centro de Investigación en Tecnologías de la Información y Comunicaciones de dicha universidad.

El mismo fue ratificado e incorporado como proyecto aprobado en el llamado a presentación interna de la UGD de proyectos de investigación N°6 mediante la Resolución Rectoral 24/A/15.

Entre las líneas con mayores resultados dentro del proyecto referido, se encuentran las de: (i) “Modelo de Simulación Dinámico de Gestión de Proyectos de Desarrollo de Software que utilizan Programación Extrema”, (ii) “Modelo de Simulación Dinámico de Gestión de Proyectos de Desarrollo de Software Bajo Scrum” y (iii) “Modelo de Simulación Dinámico de Gestión de Proyectos de Desarrollo de Software desarrollados con Crystal”. Este artículo se enfoca en la presentación de una nueva

[†] Mediante Res. Rectoral UGD N° 04/I/12.-

línea para estudiar la incorporación de prácticas de Test Driven Development a los proyectos de software a través de un modelo dinámico de simulación.

Introducción

El proceso involucrado en el desarrollo de software es complejo e implica grandes retos para el equipo de trabajo. Uno de los desafíos más grandes es garantizar la calidad del producto y mejorar la productividad a lo largo del ciclo de vida del mismo, desde el comienzo del desarrollo hasta las etapas de mantenimiento [1].

Es difícil poder determinar las características de un producto de software al inicio de su producción. El producto final se va conformando a lo largo de su propio ciclo de vida, este va sufriendo modificaciones constantes, adecuándose a las necesidades actuales del cliente.

Estos continuos cambios de requisitos, incluso en etapas tardías del desarrollo, requieren modelado adicional [2]. Causando la incorporación de código nuevo, lo cual tiende potencialmente a introducir errores, ocasionando que funcionalidades ya implantadas comiencen a fallar. Esto provoca un aumento de costo y tiempo, lo cual es uno de los dilemas principales al momento de encarar un proyecto de software.

Actualmente una de las soluciones a estos problemas es la utilización de metodologías ágiles, que incorporen herramientas que agilicen la ejecución de grandes cantidades de casos de pruebas en cuestión de minutos.

Test Driven Development (TDD) es una práctica iterativa de diseño de software orientada a objetos, que fue presentada por Kent Beck y Ward Cunningham como parte de XP [3]. Ha tenido una gran difusión en los últimos años, debido a que su objetivo es obtener código limpio que trabaje.

Esta es una técnica de programación que consiste en guiar el diseño de una aplicación por medio de pruebas unitarias. Cambia el orden tradicionalmente establecido, de manera

que primero se definen las pruebas y a partir de estas se va desarrollando la funcionalidad; repitiendo el ciclo, de acuerdo a lo que se espera que haga el software, por medio de integraciones y refactorizaciones [4].

Al utilizar la técnica TDD en lugar de realizar complejos diseños y desarrollos, se incorpora solamente las funciones que el cliente necesita evitando errores de complejidad. Sin embargo, la utilización de dicha técnica puede traer beneficios en cuanto a costos y tiempo dependiendo del tipo y tamaño de proyecto de software a desarrollar.

Se han realizado una gran cantidad de trabajos evaluando la eficacia de TDD, efectuando un análisis económico de costo-beneficio sobre la utilidad de esta práctica. En algunos de los estudios se ha llegado a la conclusión que TDD suele ser más beneficiosa para ciertos proyectos, de acuerdo a su tipo, tamaño y complejidad.

Al día de hoy no existe una métrica para determinar cuándo es conveniente utilizar esta técnica. Es por ello que la elaboración e implementación de un modelo que permita simular el desarrollo de un proyecto de software bajo la técnica de programación TDD, proporcionaría una herramienta informática fundamental a los desarrolladores de software a la hora de definir si adoptar o no este método de desarrollo.

Actualmente existen diferentes trabajos y herramientas que permiten simular la administración de proyectos de software, entre ellos se puede citar a los siguientes: “Dynamics of Agile Software Development” [1], “Modeling Agile Software Maintenance Process Using Analytical Theory of Project Investment” [2], “Modelo Dinámico de Simulación de Proyectos de Software con XP” [3] y [4], “Agile Project Dynamics: A Strategic Project Management Approach to the Study of Large-Scale Software Development Using System Dynamics” [5].

Estas herramientas presentan modelos basados en diferentes metodologías ágiles, en los cuales son

propuestos y evaluados diferentes escenarios, algunos de ellos genéricos. No obstante, en ninguno de tales trabajos se presenta un estudio que contemple las prácticas y variables particulares de TDD.

Objetivos Propuestos

Teniendo en cuenta lo planteado en la introducción en este trabajo se presentan los primeros avances en el diseño de un Modelo de Simulación Dinámico basado el modelado sistémico que de soporte a la gestión de procesos de desarrollo de software utilizan prácticas de TDD.

El modelo permite, que los administradores de proyectos, puedan evaluar qué impacto tendrán sus decisiones de gestión sobre el mismo, a lo largo del desarrollo. De esta manera el modelo refleja el efecto del uso de prácticas de TDD en proyectos de desarrollo de software. Formalmente los objetivos de la línea presentada son los siguientes:

Como objetivo general se plantea: “Construir un modelo dinámico que permita simular el desarrollo de un proyecto de software bajo la práctica Test-Driven Development para determinar la eficacia de su utilización de acuerdo a las características del proyecto”.

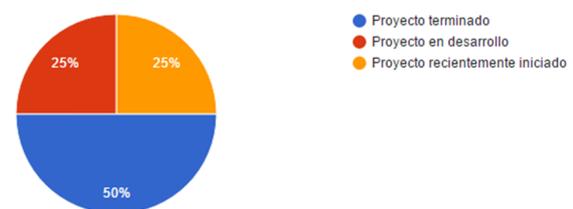
Como objetivos específicos se plantean los siguientes: A) Estudiar el funcionamiento y aplicación del método Test Driven Development. B) Analizar la bibliografía referente a la técnica Dinámica de Sistemas. C) Determinar las variables endógenas y exógenas que intervengan en el proceso de TDD. D) Modelar la estructura del sistema y su comportamiento mediante dinámica de sistemas. E) Verificar y validar el modelo con los datos provenientes de investigaciones y proyectos desarrollados con TDD. F) Evaluar la conveniencia de la adopción de TDD, de acuerdo a los distintos tipos de proyectos, según los resultados obtenidos.

Construcción del Modelo

Primeramente se realizó una encuesta para conocer características de proyectos desarrollados con Test Driven Development. Esta encuesta fue publicada en grupos o comunidades de desarrolladores que trabajan con la metodología, tanto para los de habla hispana como inglesa. Se consultó sobre cuestiones como: duración del proyecto, tamaño del equipo, grado de avance del proyecto, nivel de experiencia en TDD de los desarrolladores, etc.

En la Figura 1 se muestran algunos resultados

Grado de desarrollo del proyecto



Duración del proyecto (promedio)

18 meses

Porcentaje de uso de TDD en el proyecto

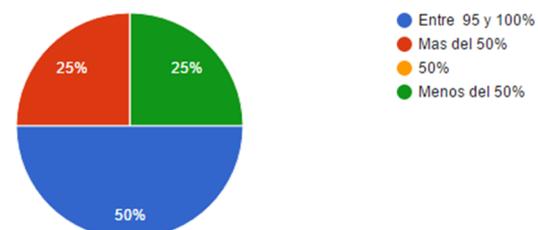


Figura 1: Algunos resultados de la encuesta realizada

En el desarrollo y construcción del modelo se utilizó la herramienta CASE VenSim PLE 5.4c (Versión Académica) [6] seguido las etapas de la Metodología de Dinámica de Sistemas [7], de esta forma, en la Fase de Conceptualización se construya el Diagrama Causal, incluyendo las variables que representan un proyecto de software Scrum y sus interrelaciones. Luego, en la Fase de Formulación se traduce el Diagrama referido a un Diagrama de Forrester, el cual ha sido dividido en Subsistemas Conservativos de

acuerdo con [8]. Finalmente, en la Fase de Evaluación se realizan las corridas de validación y experimentales del modelo.

A continuación se describe uno de los subsistemas principales del modelo, el denominado Subsistema de planificación de pruebas de aceptación. Su diagrama de influencias se puede ver en la Figura 1.

Este subsistema representa el proceso de planificación de cada una de las pruebas de aceptación lo cual es el primer paso para el desarrollo de software a través de esta metodología.

Una prueba de aceptación es la descripción del escenario desde el lado del usuario, que establece como debe funcionar y que debe realizar el sistema, módulo o submódulo, dependiendo de la generalidad de esta.

Las variable exógena que aparecen en este subsistema son Funcionalidades planificadas por iteración y Cantidad de iteraciones, cuanto mayor sea el número de estas variables mayor será el flujo de funcionalidades planificadas. Sin embargo la variable Tiempo promedio de planificación por PA tiene una relación negativa con el Flujo de funcionalidades planificadas. Al aumentar la variable endógena Funcionalidades a desarrollar, y la variable exógena Cantidad promedio de PA a desarrollar por funcionalidad, aumenta el Flujo de planificación por PA y en consecuencia la cantidad de PA a desarrollar, sin embargo este flujo se ve condicionado por el Tiempo promedio de planificación por PA.

La salida de este subsistema es la cantidad de PA a desarrollar.

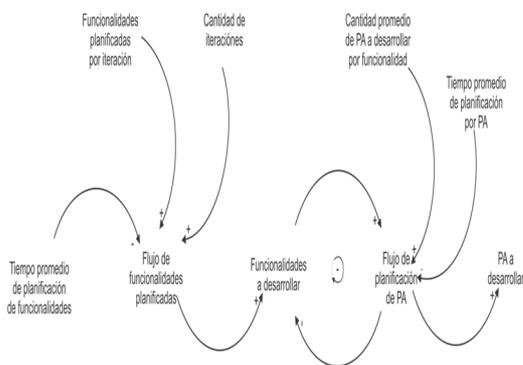


Figura 2: Subsistema de planificación de pruebas de aceptación

Resultados

En este trabajo se presentan avances en la construcción de un “Modelo de Simulación Dinámico para estudiar la incorporación de prácticas de Test Driven Development a los proyectos de software”. Al enfocarse en las particularidades de TDD, este trabajo se diferencia de otros en los que se modelan proyectos con Metodologías Tradicionales y en los que se modelan las generalidades de las Metodologías Agiles [1] [2], o bien se modela la especificidad de una de éstas últimas como en [3], [4], [9] [10] [11] [12] y [13].

Por otro lado la realización de una encuesta a desarrolladores permitió entender y conocer el grado de uso de las prácticas de TDD en los proyectos.

Como trabajo futuro se espera avanzar, con la construcción de los demás subsistemas restantes, como por ejemplo: el de desarrollo e implementación de Pruebas de Aceptación, Subsistema de planificación y desarrollo de Pruebas Unitarias, Subsistema de desarrollo de código entre otros. Otra actividad prevista es la de compilar datos de proyectos de desarrollo de software en donde se utilizó TDD, con el fin obtener datos de calidad y cantidad proyectos que siguen estas prácticas. Adicionalmente, se prevé comparar la utilización de los simuladores de modelos para Scrum [10] [11] [9] [12] y en XP [3] y [4], desarrollados anteriormente.

Formación de Recursos Humanos

El equipo de trabajo se encuentra formado por cuatro investigadores con distintos niveles de posgrado, un Doctor en Ciencias Informáticas, un Doctorando en Tecnologías de la Información y Comunicaciones, Magister y Especialista en Ingeniería de Software, un Maestrando de Ingeniería de la Web, Un maestrando en Ingeniería de Software, un investigador con nivel de grado, y seis estudiantes en período de realización de trabajos finales

de grado en el contexto de las carreras de Lic. en Sistemas de Información y de Ingeniería en Informática de la UGD. Actualmente, el número de tesinas de grado aprobadas en el contexto de este proyecto, es de tres, y otras dos en proceso de desarrollo. El número de tesis de maestría terminadas relacionadas con este proyecto es de una.

Bibliografía

- [1] Kim E. Van, Kishore Sengupta, and Luk N. Van., "Dynamics of Agile Software Development," *International Conference of the System Dynamics Society*, 2009.
- [2] Konga, Li, Liu y Chen, Jing Xiaoying, "Modeling Agile Software Maintenance Process Using Analytical Theory of Project Investment," *International Conference on Advances in Engineering 2011*, 2011.
- [3] Tamara Kasiak y Godoy Diego Alberto, "Simulación de Proyectos de Software desarrollados con XP," *XIV Workshop de Investigadores en Ciencias de la Computación.*, 2012.
- [4] Godoy Diego Alberto y Kasiak Tamara., "Modelo dinámico de simulación para la gestión de proyectos de software desarrollados con XP," in *Actas XVIII Congreso Argentino de Ciencias de la Computación*, 2012, p. 10.
- [5] Firas Glaiel, "Agile Project Dynamics: A Strategic Project Management Approach to the Study of Large-Scale Software Development Using System Dynamics," Massachusetts Institute of Technology , Tesis de Máster 2012.
- [6] Ventana System Inc. (2013) Vensim. [Online]. <http://www.vensim.com>
- [7] J Aracil, *Dinámica de Sistemas*. Madrid, España: Alianza Editorial, 1997.
- [8] Torrealdea J, *Dinámica de Sistemas. Elementos y Estructuras de un Modelo. Construyendo modelos.*: Ciencias de la Computación e Inteligencia Artificial. Universidad del País Vasco, 2003.
- [9] Diego Alberto Godoy, "Diseño de un Simulador Dinámico de Proyectos de Desarrollo de Software que utilizan Metodología Scrum," UNLP, La Plata, Tesis de Maestría 2014.
- [10] Diego Alberto Godoy, Edgardo A. Belloni, Henry Kotynski, Hector H Dos Santos, and Eduardo Omar Sosa, "Simulando Proyectos de Desarrollo de Software Administrados con Scrum," in *XVI Workshop de Investigadores en Ciencias de la Computación RedUNCI*, Ushuaia, 2014.
- [11] Diego Alberto Godoy, Eduardo Omar Sosa, Edgardo Belloni, and Henry Kotinski, "Simulación Dinámica de Gestión de Tareas en Proyectos Desarrollados Con Scrum," in *II Congreso Nacional de ingeniería informatica/ingeniería de sistemas (CoNaIISI)*, Universidad Nacional de San Luis, San Luis, 2014, 2014.
- [12] Diego Alberto Godoy, Cristian Henry Kotyński, Edgardo Anibal Belloni, and Eduardo Omar Sosa, "Un Modelo de Simulación de Proyectos scrum con dinámica de sistemas.," in *WWW/INTERNET 2015 e COMPUTAÇÃO APLICADA 2015 FLORIANÓPOLIS*, Florianópolis, 2015, Organizada por IADIS International Association for Development of the Information Society.
- [13] Cristian Henry Kotyński, Diego Alberto Godoy, Edargardo Anibal Belloni, Eduardo Omar Sosa, and Matías Anibal Rodriguez, "Análisis de la Dinámica de los Recursos Humanos en proyectos Desarrollados Con Scrum mediante Dinámica de Sistemas.," , Ciudad Autónoma de Buenos Aires, 2015.

Análisis de Foros de Discusión para la recuperación de información

Nadina Martinez Carod, Gabriela Aranda, Alejandra Cechich,
Valeria Zoratto, Pamela Faraci, Carina Noda, Mauro Sagripanti

Grupo de Investigación en Ingeniería de Software del Comahue (GIISCO)

<http://giisco.uncoma.edu.ar>

Facultad de Informática. Universidad Nacional del Comahue

Contacto: {nadina.martinez, gabriela.aranda, alejandra.cechich}@fi.uncoma.edu.ar

Resumen

La comunidad informática en general suele utilizar herramientas disponibles en la Web de soporte grupal, tanto para solucionar problemas como para el aprendizaje de alguna tarea particular. Es por ello que este tipo de herramientas de soporte han tenido un gran auge en las últimas décadas, dentro de las cuales los foros de discusión se han consolidado por ser la más utilizada.

El objetivo de los foros de discusión es el de compartir experiencias y soluciones a problemas de tópicos diversos. También funcionan como una importante fuente de información a la hora de realizar consultas y búsquedas de problemas particulares. De esta manera, cuando un usuario tiene una dificultad, realiza una pregunta sobre su problema, que suele ser respondida por diferentes usuarios, quienes proponen diversas soluciones, todo ello realizado en forma asíncrona. Bajo este prisma, los foros de discusión técnicos actúan como plataformas colaborativas para compartir soluciones que luego pueden ser reutilizadas en situaciones similares por otras personas.

En los foros de discusión los usuarios debaten un tema en común libre e informal, formando así una comunidad

respecto a un interés común. Sin embargo no es necesario ser miembro de dicha comunidad para acceder a la información conformada por las preguntas y respuestas de los ítems de discusión, lo que permite libertad de acceso a la información a cualquier usuario externo.

La consulta en foros de discusión disponibles en la Web es una tarea sencilla, pero clasificar la información obtenida no lo es, ya que en la mayoría de los casos la cantidad de información a procesar es muy grande, y muchas veces está desordenada y repetida.

Surge así el objetivo de nuestro proyecto, el cual es facilitar la tarea de un motor de búsqueda para un problema particular de un usuario desarrollando técnicas y herramientas para mejorar el reuso basado en foros de discusión.

Contexto

Nuestra línea de investigación es un proyecto que se encuentra inmerso dentro del programa “Desarrollo Orientado a Reuso”, de la Universidad Nacional del Comahue, Periodo 2013-2016. Dicho programa cuenta con tres líneas de investigación. En particular, esta línea

mantiene un acuerdo de cooperación con el Grupo Alarcos de la Escuela Superior de Informática, Universidad de Castilla-La Mancha, Ciudad Real, España.

Introducción

La Web actual se ha convertido en una plataforma que posibilita el encuentro de ideas y favorece la creación de debates científicos en foros de discusión, chats, blogs, etc., que son en cierta forma, una nueva y exhaustiva forma de revisión realizada por toda la comunidad participante. Se puede destacar, entre las herramientas mencionadas, los foros de discusión, utilizados tanto para compartir conocimiento creado por una comunidad de aprendizaje como para identificar y opinar sobre un problema particular. La naturaleza de los foros de discusión maximiza el aprovechamiento de información cuando un tema se repite o un problema surge nuevamente.

Cuando a una persona le surge un problema, ésta trata de verificar si ese mismo inconveniente o uno similar ha sucedido previamente. Para ello busca consultas realizadas previamente sobre el problema en cuestión. En un foro de discusión, cada problema o consulta, se presenta como una pregunta principal. Luego, las respuestas de diferentes participantes a esa pregunta van formando un hilo (thread) de discusión que pueden ser reutilizadas, ya sea por las mismas o por otras personas, en situaciones similares.

Las preguntas generadas por los usuarios del foro de discusión, así como las respuestas que se generan a partir de esa pregunta se suceden hasta obtener alguna solución al problema planteado.

Este intercambio de mensajes queda disponible no sólo para la comunidad del foro sino para el público en general, y puede ser reutilizado por ellos en problemas similares.

El proyecto se enfoca en hilos de foros existentes, haciendo uso de la información disponible en la Web. A partir de un análisis de calidad de la información existente en los foros, se clasifican los hilos de discusión de acuerdo a un orden de prioridad, definido mediante métricas, favoreciendo el reuso de dicha información. Esto se realiza definiendo cuando a una persona le surge un problema, ésta trata de verificar si ese mismo inconveniente o uno similar ha sucedido previamente. Para ello busca consultas realizadas previamente sobre el problema en cuestión. En un foro de discusión, cada problema o consulta, se presenta como una pregunta principal. Luego, las respuestas de diferentes participantes a esa pregunta van formando un hilo (thread) de discusión que pueden ser reutilizadas, ya sea por las mismas o por otras personas, en situaciones similares.

Si bien existen recomendadores que analizan automáticamente los hilos de un foro de discusión, como el recomendador de Chen and Persen [5], al igual que en Helic y Scerbakov [6], ambos se enfocan en un dominio de aprendizaje colaborativo. A diferencia de estas propuestas, nuestro recomendador apunta a un contexto más amplio, involucrando usuarios con distintos niveles de conocimientos del tema en cuestión. Tanto en estos trabajos como en [8] el foro utilizado es único, en cambio nuestra propuesta se enfoca en el análisis de un conjunto de foros, por lo cual se hace más

complejo con la utilización de diversos formatos de foros de discusión.

Líneas de Investigación, Desarrollo e Innovación

Este proyecto de investigación está desarrollado por el grupo de investigación de Ingeniería de Software (GIISCO), creado en el año 2002. GIISCO está compuesto por docentes y estudiantes de la facultad de Informática de la Universidad Nacional del Comahue con colaboración de otras universidades externas, nacionales y extranjeras, y su misión es el desarrollo conjunto y cooperativo de la investigación en el ámbito de la UNCo. Además se cuenta con una asesora perteneciente a la Escuela Superior de Informática, Universidad de Castilla-La Mancha, Ciudad Real, España.

El trabajo está enmarcado en el subproyecto de investigación “Reuso de Conocimiento en Foros Técnicos”, que a su vez está incluido en el Programa de Investigación “Desarrollo de Software Basado en Reuso”. Dicho programa está basado en el reuso como una estrategia de desarrollo para dar respuesta a las demandas de un menor costo tanto en la producción y mantenimiento como en la calidad del software y consta de tres líneas que convergen en el tratamiento del desarrollo de software basado en reuso. Las líneas incluidas son: Reuso Orientado al Dominio y Reuso Orientado a Servicios.

El proyecto es un trabajo colaborativo interdepartamental, en el cual están incluidos docentes del departamento de Ingeniería de Software, del Departamento

de Programación y de Teoría de la Computación. Si bien el proyecto pertenece al departamento de Ingeniería de Software, éste analiza foros de discusión con la temática específica de lenguajes de programación, incorporando este área. Sumado a esto el proyecto incluye la utilización tanto de lenguaje natural, de aprendizaje, de sentiment analysis y aprendizaje colaborativo, por lo cual se complementa con asesoría de docentes del Departamento de Teoría de la Computación.

En particular la línea de nuestro proyecto es “Reuso de Conocimiento en Foros de Discusión Técnicos” la cual por un lado determina un modelo de calidad de utilización del conocimiento existente en los foros de discusión técnicos, y por el otro define estrategias para obtener y mantener, manejar y mejorar el conocimiento existente previamente en los foros de discusión”.

Resultados y Objetivos

La problemática del proyecto de investigación, los objetivos y sus características se presentan en [3] e incluye la definición de un modelo de calidad y de gestión de conocimiento a partir de información contenida en foros de discusión técnicos. La mejora al modelo se introduce en [1] [2], lo que realiza basado a modelos de datos, de información, y de calidad de datos software [4] [7] [14].

El proceso al cual apunta nuestra propuesta se describe en [1], donde se compara la captura y el análisis de hilos de discusión, entre el buscador especializado y un navegador multipropósito estándar.

Los resultados de la encuesta a usuarios de foros técnicos, sobre su percepción en la correctitud de las soluciones sugeridas en dichos foros, se ha publicado en [9].

En [2] se presenta la herramienta para el procesamiento automático de hilos así como un caso de estudio sobre un problema particular, donde se compararon resultados obtenidos a partir de una cadena de búsqueda en un buscador de un foro de discusión específico.

En [15] se analizan distintas estrategias para clasificar hilos de discusión. Para ello se han utilizado hilos de discusión reales, recuperados del foro de discusión Stack Overflow, referidos a problemas sobre el uso del lenguaje de programación Java y, como conjunto de documentos de referencia, un repositorio de especificación de las clases Java de Oracle. El texto recuperado ha sido indexado, para determinar la relación entre los hilos y los documentos Oracle de las clases Java, con una herramienta API de recupero de información mediante indexación y búsquedas, de software abierto llamada Lucene [10]. Se está avanzando en las técnicas de análisis semántico (incluyendo stopwords, stemming, sinónimos), sobre los hilos de discusión, complementando con el diseño de un conjunto de métricas de calidad que ayuden a definir un orden en el grupo de las soluciones provistas en distintos foros para un problema particular.

Actualmente se está trabajando en generalizar los resultados publicados en [15], replicando el experimento con mayor cantidad de hilos y aplicando técnicas diferentes durante la fase de recuperación de información, así como la realización de nuevos casos de estudio. También se están estudiando alternativas

para extender el trabajo realizado, utilizando un conjunto de documentos de referencia considerando su grado relevancia.

Formación de recursos humanos

Forman parte del proyecto los siguientes miembros:

- Dos docentes investigadores del Departamento de Programación, con dedicación exclusiva, ambos con doctorado en Informática.
- Un docente investigador del Departamento de Programación, con una beca doctoral otorgada por el CONICET en diciembre de 2015.
- Dos docentes investigadores con dedicación simple, de los Departamentos de Ingeniería de Sistemas y de Programación.
- Dos estudiantes de Licenciatura en Ciencias de la Computación que están desarrollando sus tesis de grado dentro del mismo.
- Una docente del Departamento de Teoría de la Computación de la misma Facultad, que está desarrollando su tesis de Doctorado sobre técnicas de análisis de lenguaje natural. Dentro del proyecto, asesorando en temas de aprendizaje automático y lenguaje natural.
- Una docente investigadora externa, perteneciente al Grupo Alarcos de la Universidad de Castilla La Mancha, dicha docente

tiene su doctorado y una amplia trayectoria en gestión de conocimiento.

La conformación del equipo con docentes de distintos departamentos, sumado a la asesoría externa mencionada, permite el trabajo cooperativo de un grupo interdisciplinario. Además, la incorporación de estudiantes de la Facultad amplia los posibles tipos de desarrollo relacionados a la temática del proyecto.

Referencias

- [1] G. Aranda, N. Martínez Carod, P. Faraci, A. Cechich. *Hacia un framework de evaluación de calidad de información en foros de discusión técnicos*. ASSE 2013, 14th Argentine Symposium on Software Engineering, (JAIIO 2013), Córdoba, 2013.
- [2] G. Aranda, N. Martínez Carod, S. Roger, P. Faraci, A. Cechich, V. Zoratto. *Una herramienta para el análisis de hilos de discusión técnicos*. (CACIC 2014 XX), Buenos Aires, pp.803-812, 2014.
- [3] G. Aranda, N. Martínez Carod, A. Cechich, P. Faraci, C. Noda, M. Sagripanti. *Avances en reuso de conocimiento en foros de discusión técnicos*. (WICC 2014), Ushuaia, Tierra del Fuego, 2014
- [4] C. Calero, A. Caro, M. Piattini (2008), *An Applicable Data Quality Model for Web Portal Data Consumers*, World Wide Web, vol. 11, no. 4, pp. 465-484.
- [5] W. Chen, R. Persen (2009), "A Recommender System for Collaborative Knowledge".
- [6] D. Helic, N. Scerbakov (2003), *Reusing Discussion Forums as Learning Resources in WBT Systems*.
- [7] ISO/IEC 25012:2008, *Software product Quality Requirements and Evaluation (SQuaRE): Data quality model*. 2008.
- [8] H. Kuna, M. Rey, J. Cortes, E. Martini, L. Solonezen, R. Sueldo, *Generación de un Algoritmo de Ranking para Documentos Científicos del Área de las Ciencias de la Computación*, (CACIC 2013, XIX), Mar del Plata, pp. 787-796, 2013.
- [9] N. Martínez Carod, G. Aranda, M. Sagripanti, P. Faraci, A. Cechich. *Análisis de la información presente en foros de discusión técnicos*. CACIC 2013, X Workshop en Ingeniería del Software. Mar del Plata, pp. 847-856, .
- [10] M. McCandless, E. Hatcher, O. Gospodnetic. *Lucene in Action*, Second Edition, Manning Publications Co., ISBN 9781933988177, USA, 2010.
- [11] M. Nicoletti, S. Schiaffino, and D. Godoy, *Mining interests for user profiling in electronic conversations*, Expert Syst. Appl., vol. 40, pp. 638-645, Feb. 2013.
- [12] Smith y Duffy (2001), *Re-using knowledge: why, what and where*. Proceedings International Conference on Engineering Design, Glasgow (2001) .
- [13] A. Tigelaar, R. Op Den Akker and D. Hiemstra, *Automatic summarisation of discussion fora*, Natural Language Engineering, ISSN 1469-8110, Vol 16, Issue 02, pp. 161-192, 2010.
- [14] R. Wang, D. M. Strong (1996), *Beyond accuracy: What data quality means to data consumers*, Journal of Management Information Systems, vol. 12, no. 4, pp. 5-33.
- [15] V. Zoratto, G. Aranda, S. Roger, A. Cechich, *Análisis de estrategias para clasificar contenidos en foros de discusión: Un caso de estudio* ASSE 2015, 16° Simposio Argentino de Ingeniería de Software- pp. 176-190.

Dispositivos Móviles: Desarrollo y Análisis de Rendimiento de Aplicaciones Multiplataforma

Pablo Thomas⁽¹⁾, Nicolás Galdamez⁽¹⁾, Lisandro Delia⁽¹⁾, Leonardo Corbalan⁽¹⁾, Patricia Pesado⁽¹⁾

⁽¹⁾Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – Universidad Nacional de La Plata

50 y 120 La Plata Buenos Aires

{pthomas, ngaldamez, ldelia, corbalan, ppesado}@lidi.info.unlp.edu.ar

Resumen

Se presenta una línea de investigación y desarrollo, que tiene por objeto estudiar temas relacionados con aspectos de Ingeniería de Software, orientados al desarrollo de aplicaciones móviles multiplataforma.

Palabras claves: Dispositivo Móvil - Plataformas para Dispositivos Móviles - Aplicaciones Nativas- Aplicaciones Híbridas – Aplicaciones Interpretadas – Aplicaciones por Compilación Cruzada

Contexto

Esta línea de Investigación forma parte del Proyecto (2014-2018) “*Tecnologías para Sistemas de Software Distribuidos. Calidad en Sistemas y Procesos. Escenarios Educativos mediados por TICS*”, en particular del subproyecto *Ingeniería de Software en el Desarrollo de Aplicaciones para Dispositivos Móviles* del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por el Ministerio de Educación de la Nación.

Hay cooperación con Universidades de Argentina y se está trabajando con Universidades de Europa en proyectos financiados por el Ministerio de Ciencia y Tecnología de España y la AECID.

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática.

Por otra parte, se tiene financiamiento de Telefónica de Argentina en Becas de grado y posgrado.

Introducción

La computación móvil puede definirse como un entorno de cómputo con movilidad física. Un usuario debe ser capaz de acceder a datos, información u otros objetos lógicos desde cualquier dispositivo en cualquier red, al mismo tiempo que va moviéndose, cambiando su locación geográfica.

El desarrollo de aplicaciones para dispositivos móviles plantea nuevos desafíos originados en las características únicas de esta actividad. La necesidad de tratar con diversos estándares, protocolos y tecnologías de red; las capacidades limitadas, aunque en constante crecimiento, de los dispositivos; las restricciones de tiempo impuestas por un mercado altamente dinámico y la existencia de distintas plataformas de hardware y software son sólo algunas de las dificultades a las que se enfrentan los desarrolladores en esta área.

Las aplicaciones se generan en un entorno dinámico e incierto. En su mayoría se trata de aplicaciones pequeñas, no críticas, destinadas a un gran número de usuarios finales que son liberadas en versiones rápidas para poder satisfacer las demandas del mercado.

Todas las particularidades previamente mencionadas hacen que el desarrollo de software para dispositivos móviles difiera considerablemente del tradicional. Ello conduce a nuevas prácticas y metodologías que promueven el crecimiento de la Ingeniería de Software como disciplina, acompañando este proceso de desarrollo tecnológico.

La existencia de una enorme competencia en el mercado de las aplicaciones móviles, promovida en parte por la proliferación de dispositivos móviles y el número de negocios que están migrando sus servicios a Internet, revela la importancia que implica para las empresas el posicionamiento en este mercado. Para conseguirlo, es necesario reducir al máximo el tiempo de desarrollo de las aplicaciones y al mismo tiempo hacer que éstas se ejecuten en el mayor número de dispositivos posible.

Este propósito se ve obstaculizado por la excesiva fragmentación de hardware y software existente, originada por el alto número de dispositivos distintos, con sus propios sistemas operativos y plataformas de desarrollo.

La alternativa más costosa para hacer frente a esta problemática consiste en el desarrollo nativo de las aplicaciones en cada una de las plataformas existentes, utilizando el entorno de desarrollo integrado (IDE por sus siglas en inglés), el lenguaje de programación y las herramientas propias de cada plataforma.

Las aplicaciones nativas así desarrolladas poseen un conjunto de características ventajosas entre las que sobresalen el acceso a todas las capacidades del dispositivo (cámara, GPS, acelerómetro y agenda, entre otras), el alto rendimiento, la posibilidad de trabajar sin acceso a Internet y de correr en segundo plano notificando al usuario sólo en caso de requerir su atención. Estas aplicaciones pueden distribuirse a través de las tiendas en línea correspondientes. Sin embargo, el precio de todas estas ventajas es alto: no es posible reusar el código fuente entre plataformas diferentes, el esfuerzo se multiplica y se elevan los costos de desarrollo, actualización y distribución de nuevas versiones.

El desarrollo multiplataforma se contrapone al nativo y se centra en la reutilización de código. Se procura entonces optimizar la relación costo/beneficio compartiendo la misma codificación entre las versiones para las distintas plataformas. La construcción de aplicaciones Web Móviles constituye un ejemplo representativo de este enfoque. Estas aplicaciones se diseñan para correr dentro de

un navegador, se desarrollan con tecnología web bien conocidas (HTML, CSS y JavaScript), no necesitan adecuarse a ningún entorno operativo; su puesta en marcha es rápida y sencilla.

Las desventajas de las aplicaciones Web Móviles recaen sobre su rendimiento. Los tiempos de respuesta se dilatan afectados por la interacción cliente-servidor y las restricciones de seguridad impuestas a la ejecución de código por medio del navegador limitan el acceso a todas las capacidades del dispositivo. Además, al no poseer el *look and feel* de las aplicaciones nativas, resultan menos atractivas para el usuario final.

Las aplicaciones híbridas constituyen otro tipo de desarrollo multiplataforma basado en tecnologías web (HTML, Javascript y CSS) pero que, a diferencia de las anteriores, no son ejecutadas por un navegador. En su lugar, corren en un contenedor web especial con mayor acceso a las capacidades del dispositivo a través de una API específica.

Las aplicaciones híbridas permiten la reutilización de código en las distintas plataformas, el acceso al hardware del dispositivo, y la distribución a través de las tiendas de aplicaciones. Sin embargo conservan algunas de las desventajas de las aplicaciones Web Móviles: la utilización de componentes no nativos en la interfaz perjudica la experiencia de usuario, y la ejecución se ve ralentizada por la carga asociada al contenedor web.

Otro tipo de aplicación multiplataforma lo constituyen las aplicaciones interpretadas, las cuales son traducidas en su mayor parte a código nativo, mientras que un resto se interpreta en ejecución. Se implementan de forma independiente de las plataformas utilizando diversas tecnologías y lenguajes, tales como Java, Ruby y XML, entre otros.

La obtención de interfaces nativas constituye una de las principales ventajas de este tipo de aplicaciones, y la dependencia total con el entorno de desarrollo el obstáculo más notable. Appcelerator Titanium es el entorno de desarrollo más popular.

Finalmente, las aplicaciones generadas por compilación cruzada también constituyen un tipo de desarrollo multiplataforma. Estas aplicaciones se compilan de manera nativa creando una versión específica de alto rendimiento para cada plataforma destino. Ejemplos de entornos de desarrollo para generar aplicaciones por compilación cruzada son Applause, Embarcadero Delphi XE6 y Xamarin.

Líneas de Investigación y Desarrollo

- Tipos de Aplicaciones para Dispositivos Móviles
- Metodologías y Técnicas de la Ingeniería de Software y su aplicación en el desarrollo de software para dispositivos móviles.
- Aplicaciones Web Móviles
- Aplicaciones Móviles Híbridas (PhoneGap, Sencha Touch)
- Aplicaciones Móviles Interpretadas (Appcelerator Titanium)
- Aplicaciones Móviles generadas por compilación cruzada (Xamarin, Applause, Embarcadero Delphi XE6)
- Análisis y estudio comparativo de los aspectos inherentes al desarrollo de aplicaciones móviles multiplataforma para los cuatro tipos anteriormente referidos.
- Análisis y estudio comparativo de rendimiento sobre aplicaciones móviles multiplataforma.

Resultados esperados/obtenidos

Los resultados esperados/obtenidos se pueden resumir en:

- Avanzar en la capacitación continua de los miembros involucrados en esta línea de investigación.
- Utilizar y comparar el uso de diversos frameworks para el desarrollo de aplicaciones móviles nativas y multiplataforma.

- Estudiar las estrategias de desarrollo multiplataforma, aplicadas a la implementación de una aplicación móvil de apoyo al docente en el aula, cuyo objetivo principal es la toma de asistencia *in situ* vía bluetooth.
- Analizar y comparar el impacto que distintos frameworks de desarrollo multiplataforma tienen sobre el rendimiento de las aplicaciones móviles construidas con ellos. Enfocar este análisis sobre la velocidad de procesamiento, el acceso a disco, el consumo de batería y otras capacidades; considerando además el tipo, gama y sistema operativo del dispositivo.
- Cuantificar las diferencias de rendimiento entre las aplicaciones multiplataforma y las aplicaciones nativas.
- Se ha realizado el estudio comparativo entre diferentes tipos de desarrollo de aplicaciones móviles a partir de la implementación de diversas aplicaciones con idéntica funcionalidad, que permiten interactuar con la plataforma de e-learning Web-UNLP. Esto incluye el desarrollo de:
 - Una aplicación nativa para Android y otra para iOS.
 - Una aplicación multiplataforma Web Móvil.
 - Dos aplicaciones multiplataforma híbridas, una de ellas utilizando PhoneGap con JQuery Mobile y la otra desarrollada con el framework Sencha Touch
 - Una aplicación interpretada con Appcelerator Titanium 3
 - Dos aplicaciones generadas por compilación cruzada, una de ellas desarrollada con Embarcadero Delphi XE6 y la otra con Xamarin/Visual Studio.
- En relación con el análisis de rendimiento, se ha realizado un conjunto de pruebas focalizadas en la velocidad de procesamiento. Ello ha permitido efectuar

comparaciones de tiempos de ejecución en aplicaciones que demandan cálculo intensivo. Los escenarios de prueba diseñados incluyeron los sistemas operativos Android y iOS ejecutándose sobre diversos dispositivos smartphones y tablets. Se ha medido el impacto que los frameworks de desarrollo Apache Cordova, Titanium, y Xamarin (desarrollo híbrido, aplicación interpretada y generada por compilación cruzada respectivamente) tienen sobre el rendimiento de las aplicaciones con ellos desarrolladas.

Formación de Recursos Humanos

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con el proyecto. Además participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

Referencias

1. Hayes, I. S. *Just Enough Wireless Computing*. Prentice Hall Professional Technical Reference. 2002. ISBN:0130994618
2. Spyros Xanthopoulos, Stelios Xinogalos, *A Comparative Analysis of Cross-platform Development Approaches for Mobile Applications*, BCP 2013, Greece
3. Tracy, K.W., *Mobile Application Development Experiences on Apple's iOS and Android OS*, Potentials, IEEE, 2012.
4. Delia L., Galdamez N., Thomas P, Pesado P., *Un Análisis Experimental de Tipo de Aplicaciones para Dispositivos Móviles*, CACIC 2013, CAECE Mar del Plata, Octubre 2013
5. <http://www.appcclerator.com/>
6. <https://github.com/applause/applause>
7. <https://www.embarcadero.com/es/products/delphi>
8. <http://xamarin.com/>
9. Abrahamsson, P. *Mobile software development -the business opportunity of today*. Proceedings of the International Conference on Software Development, (pp. 20-23). 2005. Reykjavik.
10. <http://devgirl.org/2012/12/04/easy-phonegap-push-notifications-with-pushwoosh/>
11. <http://phonegap.com/>
12. <http://jquerymobile.com/>
13. <http://backbonejs.org/>
14. Digital Possibilities. Mobile Development Frameworks Overview <http://digital-possibilities.com/mobile-development-frameworks-overview/>
15. Markus Falk. Mobile Frameworks Comparison Chart, <http://www.markus-falk.com/mobile-frameworks-comparison-chart/>
16. Anup Kumar y Bin Xie, *Handbook of Mobile Systems Applications and Services*. Editorial CRS Press, ISBN 978-1-4398-0152-9, Año 2012.
17. Sambasivan, D.; John, N.; Udayakumar, S.; Gupta, R., *Generic framework for mobile application development*, Internet (AH-ICI), 2011 Second Asian Himalayas International Conference on Computing & Processing (Hardware/Software).
18. Choi, Y.; Yang, J.-S.; Jeong, J., *Application framework for multi platform mobile application software development*, Advanced Communication Technology, 2009. ICACT 2009. 11th International Conference on Computing & Processing (Hardware/Software)
19. Anthony Wasserman , Carnegie Mellon Silicon Valley, *Software Engineering Issues for Mobile Application Development*, 2º Workshop on Software Engineering for Mobile Application Development, MobiCASE '11, Santa Monica, California, USA, October 2011.
20. Delia L., Galdamez N., Corbalan L., Thomas P, Pesado P., *Multi-Platform Mobile Application Development Analysis*, IEEE Ninth International Conference on Research Challenges in Information Science, May 2015, Athens, Greece, ISBN 978-1-4673-6630-4
21. Delia L., Galdamez N., Corbalan L., Thomas P, Pesado P., *Un Análisis comparativo de rendimiento en Aplicaciones Móviles Multiplataforma*, CACIC 2015, UNNOBA Junín, Octubre 2015.

Experiencias en el desarrollo de Sistemas de Software Distribuidos.

Patricia Pesado⁽¹⁾, Rodolfo Bertone⁽¹⁾, Pablo Thomas⁽¹⁾, Luciano Marrero⁽¹⁾, Ariel Pasini⁽¹⁾, Lisandro Delia⁽¹⁾, Nicolás Galdamez⁽¹⁾, Eduardo Ibañez⁽¹⁾, Cesar Estrebou⁽¹⁾, Alejandra Ripodas⁽¹⁾, Verónica Aguirre⁽¹⁾, Germán Cáseres⁽¹⁾, Santiago Medina⁽¹⁾, Matías Dell’Oso⁽¹⁾, Juan Manuel Paniego⁽¹⁾, Martín Pi Puig⁽¹⁾

⁽¹⁾Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 120 La Plata Buenos Aires

{ppesado, pbertone, pthomas, lmarrero, apasini, ldelia, ngaldamez, eibanez, cesarest, aripodas, vaguirre, gcaseres, smedina, mdelloso, jmpaniego, mpipuig}@lidi.info.unlp.edu.ar

Resumen

Esta línea de investigación y desarrollo está centrada en los sistemas de software distribuidos; desde hace más de 15 años el III-LIDI trabaja con metodologías, técnicas y herramientas de Ingeniería de Software para sistemas distribuidos, orientadas a diferentes clases de sistemas, tales como sistemas WEB, sistemas de E-government, sistemas para dispositivos móviles.

Esta línea ha significado la transferencia de conocimiento a diferentes sectores, tanto públicos como privados o mixtos.

Palabras claves: Sistemas Distribuidos – Ingeniería de Software - Metodologías de Desarrollo – Sistemas Web – E-Government

Contexto

La línea de Investigación que se presenta en este trabajo está enmarcada en el Proyecto F016-Tecnologías para Sistemas de Software Distribuidos. Calidad en Sistemas y Procesos. Escenarios educativos mediados por TICs (2014-2017), del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por el Ministerio de Educación. En particular en el subproyecto “Métodos y Procesos para la gestión de Sistemas de Software Distribuidos. Aplicaciones”.

Introducción

El desarrollo de sistemas distribuidos, como cualquier sistema de software necesita de las actividades definidas en un ciclo de vida de software para su correcta resolución. En algunos casos las técnicas y herramientas utilizadas habitualmente, deben ser adaptadas a estos tipos de sistemas, cuya característica principal es que se ejecutan sobre un conjunto de computadoras conectadas a una red. Estos sistemas deben tener soporte de software distribuido, que permita coordinar las actividades y compartir recursos, de manera que el usuario perciba una única facilidad de cómputo integrado, aunque ésta pueda estar implementada por varias máquinas en distintas ubicaciones [1].

La Ingeniería de Software comprende la aplicación de principios científicos para generar una solución elaborada de software que resuelva un problema determinado, y el mantenimiento subsecuente de ese sistema de software hasta el final de su vida útil [2][8][9]. La utilización de estas prácticas para resolver Sistemas Distribuidos hace necesaria su adaptación en función de las características de dichos sistemas. [3]

Existen diversos modelos de procesos que permiten generar software que cumpla las expectativas del cliente y que para su desarrollo se pueda tener una agenda y costeo adecuado. Estos diferentes modelos presentan actividades rectoras que están presentes, en general, en cualquier metodología [10]:

- 1) especificación del problema (ingeniería de requerimientos)
- 2) gestión del proyecto (calendarización, análisis de riesgo, métricas y análisis de costo)
- 3) diseño e implementación,
- 4) validación
- 5) evolución del software.

Asociado con la Ingeniería de Software de Sistemas Distribuidos, está el problema de utilizar un entorno WEB para los servicios que ofrece el Sistema. El desarrollo de arquitecturas centradas en un servidor (o un conjunto de servidores distribuidos) que ofrecen una interfaz WEB a los usuarios, ha generado un importante desarrollo en la investigación de metodologías y herramientas orientadas a Sistemas WEB, así como ha generado la necesidad de establecer nuevas métricas y parámetros de aseguramiento de la Calidad para tales Sistemas [4] [5] [6][7].

En esta línea de investigación interesa particularmente la temática de E-Government. El gobierno electrónico consiste en el uso de las tecnologías de la información y el conocimiento en los procesos internos de gobierno en la entrega de los productos y servicios del Estado tanto a los ciudadanos como a la industria. Debe centrarse en la inclusión de los ciudadanos de una manera participativa a través de las tecnologías de la información y la comunicación [15] [16] [17]. Al concepto clásico de E-Government que se ha enfocado en poner los servicios tradicionales del Estado al alcance del ciudadano (Consultas en línea, Gestión de trámites, Expedientes digitales, Voto Electrónico, Consultas populares, etc) [18][19][20][26] se agrega la concepción de E-Citizen, es decir un ciudadano capacitado para interactuar con el Estado, empleando Tecnología. [21][22][23][24][25][27].

Desde el año 2003 el Instituto trabaja en aplicaciones en esta área, entre las cuales se destacan prototipos de hardware y software de distintos tipos de votaciones (urnas electrónicas, ambientes de votación, comunicaciones, entre otras)[11][12][13][14].

Un sistema de votación electrónica utiliza las TIC (Tecnologías de la Información y Comunicación) para llevar a cabo el proceso de emisión de votos, haciendo el recuento de votos simple y rápido.

Existen numerosas implementaciones que apoyan el proceso de votación, algunos de ellos destinado sólo a ciertas fases del proceso, mientras que otros a resolver todas las actividades que se llevan a cabo durante una elección. Estos últimos se denominan sistemas de voto electrónico.

Los requisitos básicos de un sistema de votación electrónica son:

- Sólo los que están en el registro electoral debe ser capaz de votar;
- Ningún votante debe ser capaz de tomar el lugar de otro, o votar más de una vez;
- El voto de ser secreto, la persona debe emitir su voto y éste no debe ser identificable;
- Los votos en blanco deben ser posibles.

Una vez que el proceso electoral está habilitado por el consejo de autoridades de la Junta Electoral, el proceso de votación incluye tres etapas principales:

1. identificación del votante,
2. registro del voto,
3. una vez finalizado el acto electoral, realizar es escrutinio.

El software desarrollado para el voto electrónico debe ser auditable, lo que significa que las características de código abierto en un sistema operativo de software libre con la firma digital serían útiles para la fiabilidad.

Se han analizado y caracterizado modelos y tecnologías posibles, realizando varias experiencias prácticas. En particular se han analizado tres variantes: Voto electrónico presencial (VEP), Voto electrónico semipresencial (VESP) y Voto remoto (VER)

VEP: Una de las soluciones más integrales para la registración del sufragio, es la urna electrónica, un dispositivo que permite al elector visualizar las opciones de voto a través de una pantalla táctil, verificar su elección con

el voto impreso que se muestra en una ventana y confirmar el mismo para actualizar los contadores de votación y deslizar automáticamente el voto en una urna alojada en el “cajero/kiosco”. Estos sistemas son llamados Sistemas de Registro Electrónico Directo (DRE). El VEP se organiza en puestos de votación distribuidos, en los que existe un padrón de votantes en la mesa de autoridades donde se presenta físicamente el elector, conectada a la urna electrónica. Finalizada la elección se realizará el recuento automático de los votos de la urna.

VESP: Combina las características del voto presencial en cuanto a los puestos de votación donde se identifican los electores y efectúan su opción de voto; con la transmisión remota del voto realizado a una urna distante físicamente, donde se guarda el respaldo papel y se registran los resultados y sobre la que se realizará el recuento de los votos recibidos una vez finalizada la elección.

VER: Consiste en una aplicación WEB, que puede ser utilizada por el elector a través de un navegador web desde un dispositivo conectado a internet en cualquier lugar del mundo. El voto del elector es transmitido por una conexión segura a un servidor donde se almacenan los resultados. Una vez finalizada la elección, las autoridades electorales realizan el recuento a través de la aplicación web.

Líneas de Investigación y Desarrollo

Se detallan a continuación las principales líneas de investigación y desarrollo que representan una continuidad respecto de años anteriores:

- Conceptos de procesamiento distribuido. Arquitectura, comunicaciones y software. Middleware.
- Metodologías de especificación, validación y desarrollo de SSD.
- Metodologías ágiles de desarrollo utilizando frameworks propios y

disponibles de uso libre, con diferentes tecnologías.

- Evaluación de técnicas para mejorar el proceso de Ingeniería de Requerimientos.
- Lenguajes y ambientes para procesamiento distribuido.
- Reingeniería de sistemas complejos que migran por downsizing a esquemas cliente-servidor distribuidos.
- Sistemas basados en tecnologías Grid/Cloud. Enfoque a las aplicaciones de E-Citizen.
- Sistemas distribuidos para el tratamiento de información no estructurada (Big-Data). Aplicaciones.
- Voto electrónico presencial y Voto por Internet. Arquitecturas adaptadas a la legislación vigente. Inclusión de electores.
- Sistemas de identificación segura en tiempo real. Identificación biométrica.
- Seguridad y tolerancia a fallas en sistemas de E-Government.
- Certificación de software y hardware para sistemas de E-Gov. Normas.
- Nuevas tecnologías de E-Gov, en particular M-Gov (Mobile Government)

Resultados esperados/obtenidos

Algunas transferencias realizadas y en curso:

- **Voto Electrónico Presencial.** Desde 2003, se utilizan adaptaciones de la urna electrónica en las elecciones de alumnos de distintas universidades (los mismos pueden votar a “Claustro de estudiantes y Centro de estudiantes” o “Sólo Centro” de acuerdo a su condición), incluyendo alumnos no videntes. También en elecciones de graduados, de profesores, de no docentes, en congresos para votación de las mejores presentaciones o en eventos para la elección de las mejores propuestas. Está en proceso de desarrollo una nueva arquitectura de Urna Electrónica con utilización de

microcontroladores del tipo Raspberry Pi 2.

- **Voto Electrónico Semipresencial.** Desde 2007 se utiliza voto semi-presencial en elecciones estudiantiles de la UNLP en sedes regionales.
- **Voto Electrónico Remoto:** Desde 2007 se utiliza voto remoto en elecciones de distintas organizaciones públicas y privadas (unidades de investigación, consejos y cajas profesionales), para diferentes conformaciones de autoridades.
- **Sistema de Alerta temprana.** El III-LIDI está desarrollando un Sistema WEB de Alerta Temprana para la ciudad de La Plata (SATLP) para la prevención de inundaciones, en el marco del proyecto PIO CONICET-UNLP “Construcción de un sistema integrado de gestión del riesgo hídrico en la región del Gran La Plata”. El objetivo es contar con un sistema de alerta temprano que permitirá monitorear no sólo las precipitaciones sino el crecimiento de cada cuenca hidrológica de la región. El proyecto busca comunicar estaciones de censado pluviométrico con una base de datos central y que la información producida sea visible en la web desarrollada a tal efecto. Las etapas del desarrollo son:
 - 1- visualización de los datos de lluvias;
 - 2- agregar la caracterización de la severidad de la tormenta en tiempo real;
 - 3- dar resultados del simulador (modelo hidrodinámico);
 - 4- presentar un pronóstico de inundación que defina un alerta hidrológico.

Esta experiencia se está implementando para la subcuenca del Arroyo Pérez. Este proyecto tiene el potencial de poder ser utilizado en otras cuencas de la Provincia.

- **Sistema para almacenamiento y recuperación inteligente de información digitalizada.** Durante WICC 2015 se presentó este sistema en una etapa de

análisis y diseño. Actualmente está implantado en la Facultad Ciencias Económicas y en la Facultad de Informática de la UNLP.

- **Sistema de Gestión Integral para el CONICET La Plata.** Este sistema, que abarca las áreas de Recursos Humanos; Comercio Exterior, Compras y Patrimonio; Administración Contable; Tesorería; entre otras, ya se encuentra operativo.
- **Plataforma de Educación a Distancia.** El sistema WEBUNLP fue desarrollado al comienzo del milenio por el III-LIDI. Desde ese momento fue utilizado por numerosas cátedras de la Facultad de Informática, la Universidad Nacional de la Plata, y otras universidades del país. El proyecto contó con mantenimiento preventivo y correctivo, pero el transcurso del tiempo demostró que era necesario un mantenimiento perfectivo que incorporara una serie de nuevas funcionalidades al entorno. Desde 2015 el sistema está siendo redefinido.

Metas/Resultados esperados (en continuidad con el proyecto que se viene realizando) :

- Desarrollar soluciones a problemas concretos de software de sistemas distribuidos, poniendo énfasis en el desarrollo de metodologías y herramientas específicas para clases de aplicaciones.
- Analizar metodologías aplicables a Sistemas Distribuidos utilizando frameworks de desarrollo, específicos para diferentes entornos.
- Ampliar herramientas que soporten la evaluación de calidad, eficiencia y relación de esfuerzo/costo. Esta meta relacionada con el subproyecto de gestión de la calidad del Instituto.
- Continuar evolucionando los prototipos de voto electrónico presencial y remoto utilizables en diferentes modelos de elecciones/consultas.
- Transferir al sector productivo nacional.

- Formar recursos humanos de grado y postgrado en la temática.
- Conformar vínculos y acuerdos de cooperación con otras universidades del país y del exterior.
- Publicar los resultados y participar en eventos científicos/tecnológicos para compartirlos.

Formación de Recursos Humanos

Los integrantes de esta línea de investigación dirigen Tesinas de Licenciatura y Tesis de Postgrado en la Facultad de Informática, y Becarios del Instituto en temas relacionados con el proyecto. Además participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP y en otras universidades del país y del exterior.

Referencias

- [1] G. Coulouris. Distributed Systems – Concepts and Design. Addison-Wesley. 1994.
- [2] R. Pressman. Ingeniería de Software: Un Enfoque Práctico. McGraw-Hill. 2002
- [3] Pleeger. Ingeniería de Software: Teoría y Práctica. Prentice-Hall. 2002
- [4] Stephen Kan. Metrics and Models in Software Quality Engineering (2nd Edition). Addison Wesley. 2003
- [5] Offutt J., “Quality Attributes of Web Software Applications”. IEEE Software: Special, Issue on Software Engineering of Internet Software 19 (2):25-32, Marzo / Abril 2002.
- [6] Wu, Y. y Offutt, J. “Modeling and testing web-based Applications”. <https://citeseer.ist.psu.edu/551504.html>: 1-12, Julio 2004
- [7] Piattini, M; Oktaba, H; Pino, F; Orozco, M; Alquicira, C. COMPETISOFT. Mejora de Procesos Software para Pequeñas y Medianas Empresas y Proyectos. Editorial RaMa. ISBN; 978-84-7897-901-1. 2008
- [8] Ingeniería de Software, Ian Sommerville, Pearson, Addison Wesley, 2006
- [9] Software Engineering Institute. Disponible en <http://www.sei.cmu.edu/risk/>
- [10] A spiral model of software development and enhancement. B. Boehm, IEEE Computer, 21(5), 61-72. 1988
- [11] Feierherd G., De Giusti A., Pesado P., Depetris B. “Una aproximación a los requerimientos del software de voto electrónico de Argentina”. CACIC 2004.
- [12] Pesado P., Feierherd G., Pasini A. “Especificación de Requerimientos para Sistemas de Voto Electrónico”. CACIC 2005.
- [13] Pesado P., Pasini A., Ibáñez E., Galdámez N., Chichizola F., Rodríguez I., Estrebow C., De Giusti A. “E-Government- El voto electrónico sobre Internet”. CACIC 2008.
- [14] Carri J., Pasini A., Pesado P., De Giusti A. “Reconocimiento biométrico en aplicaciones de E-Government. Análisis de confiabilidad / tiempo de respuesta.” CACIC 2007.
- [15] Center for democracy and technology E-Government Handbook. 2002. <http://www.cdt.org/egov/handbook/>
- [16] Jones A., Williams L. “Public Services and ICT - FINAL REPORT. How can ICT help improve quality, choice and efficiency in public services?”. London: The Work Foundation. 2005.
- [17] London: National Audit Office. “Better Public Services through e-government”. Report HC 704-I Session 2001-2002.
- [18] Washington DC: Office of Management and Budget. “E-Government Strategy: Simplified Delivery of Services to Citizens”. OMB. 2002.
- [19] “Citizen Centric Government: Global Best Practice in Delivering Agile Public Services to Citizens and Businesses”. London: Gov3 Ltd. 2006.
- [20] Brussels: European Commission. “The Role of eGovernment for Europe's Future”. Report COM(2003) 567 Final. Europe. 2003.
- [21] Wang L., Bretschneider S., Gant J. “Evaluating Web-based e-government services with a citizen-centric approach”. Proceedings of the 38th Hawaii International Conference on System Sciences. 2005.
- [22] Transforming Public Services: The Next Phase of Reform. Edinburgh: Scottish Executive”. Scotland 2006.
- [23] Clift S. “e-Government and Democracy: Representation and Citizen Engagement in the Information Age”. 2004.
- [24] O'Donnell S., McQuillan H., Malina A. “eInclusion: expanding the Information Society in Ireland. Dublin: Government of Ireland. Information Society Commission”. 2003.
- [25] Juma C., Yee-Cheong L. “Reinventing global health: the role of science, technology and innovation”. Lancet 2005.
- [26] Laguna A., Ferri Tormo R., Hernandez V., Peñarrubia J. “gCitizen: uso de tecnologías Grid para la interoperabilidad entre Administraciones Públicas”. IX Jornadas sobre Tecnología de la Información para la Modernización de las Administraciones Públicas. Sevilla. 2006.
- [27] Brunner J. “Educación: escenarios de futuro. Nuevas Tecnologías y sociedad de la información”. PREAL, Santiago de Chile. 2000.

Guías para aplicación de Normas de Calidad para los procesos de Ingeniería de Software en productos desarrollados con Lenguajes de Programación Open Source: relevamiento y aplicación en PYMES de la zona de influencia de la UNER Concordia

Cristian D. Pacifico⁽¹⁾, Martín M. Pérez⁽¹⁾, Mónica D. Tugnarelli⁽¹⁾, Marcelo A. Falappa⁽²⁾

⁽¹⁾ Facultad de Ciencias de la Administración – Universidad Nacional de Entre Ríos

Monseñor Tavella 1424 – (CP 3200) Concordia, Entre Ríos
Teléfono: (+54) (345) 423 1400 – Fax: (+54) (345) 423 1410

⁽²⁾ Departamento de Ciencias e Ingeniería de la Computación – Universidad Nacional del Sur

Avenida Alem 1253 – (CP 8000) Bahía Blanca, Buenos Aires
Teléfono: (+54) (291) 459 5135 – Fax: (+54) (291) 459 5136

e-mails: martinmperez@gmail.com, mfalappa@cs.uns.edu.ar

Resumen

El Software Libre u *Open Source* es una forma de licenciamiento que da a los usuarios la libertad de compartir, estudiar, y modificar el software. Este paradigma ha tenido impedimentos para su adopción en ambientes empresariales y comerciales debido a concepciones erróneas. En la actualidad varias de estas ideas se desmitificaron brindando una dosis de realidad a las oportunidades y obstáculos de esta filosofía. Esto se manifiesta en una evolución y aceptación paulatina de la adopción de software libre como temática central en la Ingeniería en Software.

Este contexto plantea la necesidad de contar con estándares de medición y certificación de calidad específicos para productos de Software Libre y de las competencias y desempeño de empresas y

profesionales que utilicen, generen y promocionen el Software Libre. El presente proyecto de investigación propone la formulación de marcos metodológicos y modelos decertificación relacionados con software libre y lenguajes de programación *open source*, ya sea para empresas usuarias como desarrolladoras. De esta manera, se le dará valor agregado y ventaja competitiva a las empresas y profesionales que utilicen, desarrollen y/o promuevan el Software Libre y las tecnologías abiertas.

Palabras clave: Ingeniería de Software, Lenguajes de Programación, Software Libre, Calidad, Certificación.

Contexto

La línea de I+D de este proyecto tiene como antecedente destacado el proyecto de desarrollo tecnológico para Fortalecimiento a la Innovación Tecnológica en Aglomerados Productivos FIT-AP 003/12 “Cluster de Software Libre” y en particular la Acción 1: “CERTASOL (CEntro de Referencia en Tecnologías Abiertas y SOftware Libre)”. Esta acción es llevada a cabo bajo la responsabilidad de la Facultad de Ciencias de la Administración de Universidad Nacional de Entre Ríos. Esta unidad académica y varias empresas de desarrollo de software conforman el aglomerado productivo exclusivamente dedicado a apoyar y fomentar el desarrollo e implementación de software libre. Uno de los objetivos de este proyecto FIT-AP es establecer un centro de referencia para certificar a profesionales, software y usuarios en el uso, desarrollo y promoción de tecnologías y productos basados en Software Libre, y de un observatorio para el análisis y seguimiento de información referida a dichas tecnologías.

El Proyecto de Investigación y Desarrollo (PID) 7049 que se presenta aquí busca articular con la acción anteriormente descrita estableciéndose como ámbito científico-académico para el desarrollo y redacción de guías y normas para validación y certificación de productos y procesos de Software Libre y Tecnologías Abiertas.

Introducción

La Industria del Software es considerada de alto valor agregado y generadora de empleo, encuadrándose dentro de las llamadas empresas basadas

en el conocimiento. En Argentina se registra un alto crecimiento del sector debido a la informatización, a la modernización del sector PYMES, y a las políticas estatales favorables al sector del software. La inclusión de la cadena de valor del Software y los Servicios Informáticos en el Plan Estratégico Industrial 2020 [Presidencia, 2012] da un aspecto estratégico a la tecnología informática ya que se la define como vía de mejoras transversales en productos y procesos de todas las ramas productivas.

En el mismo sentido, el Libro Blanco de la Prospectiva TIC Proyecto 2020 del Ministerio de Ciencia, Tecnología e Innovación Productiva [Presidencia, 2009], en la discusión de aspectos de la problemática de las TICs para identificar áreas de promoción prioritarias en la Argentina, resalta que el dominio de la tecnología *open source* o de software libre debe considerarse como una potencial fortaleza a explotar por el país. Paralelamente, en los últimos años la Argentina ha aumentado la producción y exportación de software, logrando un posicionamiento distinguido en mercados internacionales por la calidad tanto del software que genera como de los profesionales en la disciplina. El aumento en la demanda de profesionales capacitados en Software Libre ha logrado ocupar cerca del 40% de ese mercado, con más de ciento sesenta empresas registradas que ofrecen aplicaciones y sistemas basados en Software Libre [CESSI, 2013].

El Software libre es una forma de licenciamiento que brinda a los usuarios la libertad de compartir, estudiar, y modificar el software ya desarrollado por otros. Uno de los principales impedimentos para la adopción e implementación de productos y servicios basados en Software Libre es la percepción, en los niveles gerenciales y

profesionales, de varias concepciones erróneas [Berg, 2013]. Dentro de estos preconceptos, se destacan la falta de información y de soporte, la supuesta hostilidad del software libre hacia la propiedad intelectual, la informalidad de los proyectos de software libre, la falta de responsabilidad y respaldo que sí se garantizaría con empresas propietarias de software, y la baja calidad y poca seguridad que encierra un producto que no tiene un mantenimiento formalizado, entre otras. Esto incrementa la sobrevaloración a favor de los productos propietarios, que se sustenta principalmente en la confianza y seguridad que aseguran brindar las empresas proveedoras de los mismos; siendo esto una de sus prioridades en las estrategias de marketing que desarrollan. “*El Software Libre, como tal, no tiene dueños y, por ende, no tiene responsables*” es una de las afirmaciones falaces en detrimento del uso de tecnologías de software libre.

En la última década se han desmitificado varias de las objeciones al Software Libre, logrando evolucionar en concepciones más realistas sobre las ventajas y debilidades del mismo en ambientes productivos y empresariales [Cooper, 2014; da Rosa, 2007; Wasserman, 2009]. El Software Libre es, sin lugar a dudas, una revolución tecnológica que ha cambiado la manera de entender las libertades de los usuarios, los procesos de desarrollo y ha creado un modelo de negocios más económico y rentable. En este sentido, miles de aplicaciones de alto nivel de desarrollo han logrado insertarse con pleno éxito en el mercado del software y varias empresas de referencia en tecnología han optado por iniciativas y proyectos de Software Libre.

Según los últimos informes publicados por The Linux Foundation [2010], se

advierte un fuerte y sostenido crecimiento a nivel internacional, en la demanda de profesionales certificados en entornos Linux y desarrollo en software libre; de igual manera, se muestra una tendencia a migrar a plataformas de Software Libre por varias razones, siendo las más importantes la superioridad técnica de los productos, el bajo costo y la seguridad. El contexto que favorece a la adopción del desarrollo de software *open source* con plataformas libres también plantea exigencias respecto a validar y certificar calidad en los productos software y procesos de ciclos de vida que involucra la ingeniería en software. La implementación de normas de calidad y su correspondiente certificación traen significativos beneficios:

- Ofrece un lenguaje común entre sector productivo, ciudadanos, comunidad y las Administraciones Públicas.
- Fortalece el desarrollo de la Industria Nacional de Tecnologías de Información Libre.
- Facilita el intercambio de información y experiencias en Tecnologías de Información Libre.
- Establece niveles de calidad y seguridad en productos y servicios inherentes al tipo de ciclo de vida involucrado.

La *certificación* es un mecanismo mediante el cual se garantiza que los productos, servicios y procesos de Tecnologías de Información cumplan con Normas Técnicas de Calidad previamente establecidas. Algunos de los beneficios de la Certificación son las siguientes:

- Contribuye a la adopción de las normas técnicas aprobadas.
- Aumenta la calidad de las Tecnologías de Información Libre utilizadas por el Estado.
- Fortalece la Industria Nacional de las Tecnologías de Información.

- Genera nuevas oportunidades de servicios.

Líneas de Investigación, Desarrollo e Innovación

Por un lado, se espera lograr una base de conocimiento acerca de las PYMES de la región que tengan procesos de ciclo de vida de software, puntualizando características de sus procesos de ciclo de vida y adopción de Software Libre. Para esto se prevé la realización de un relevamiento exhaustivo de las PYMES de la región.

Por otro lado, mediante la creación de un marco de referencia para la aplicación de normas de calidad de productos desarrollados con lenguajes de programación *open source* y/o tecnologías libres, se intentará seleccionar normas específicas de calidad para procesos y productos de Ingeniería en Software, y especificar propósito, límites, y destinatarios de los modelos y guías de aplicación de las normas de calidad elegidas.

Por último, se espera especificar y redactar guías o procedimientos para la aplicación de normas de calidad, tanto de productos como de procesos de ciclo de vida que utilicen plataformas *open-source*.

Resultados y Objetivos

Uno de los principales objetivos de este proyecto de investigación es especificar modelos y guías de certificación para productos de Software Libre y procesos de ciclo de vida que involucren plataformas *open source*.

A partir del desarrollo de un marco metodológico y de un modelo de

certificaciones específicas relacionadas al Software Libre se pretende:

- Crear un marco de referencia para la certificación de calidad de productos desarrollados con lenguajes de programación *open source* y/o tecnologías libres.
- Especificar guías o referenciales para la aplicación de normas de calidad en los procesos de ciclo de vida de proyectos de Software Libre.
- Delinear estrategias de implementación de dichos modelos en PYMES de la región.

Formación de Recursos Humanos

Se pretende realizar la formación en investigación de 1 (un) becario de Iniciación en la Investigación afectado a este proyecto PID; 2 (dos) becarios con becas de formación de recursos humanos de la UNER y se prevé el desarrollo de al menos 4 (cuatro) proyectos de Trabajo Final para la carrera Licenciatura en Sistemas.

Además, se espera que dos integrantes del equipo defiendan sus tesis de maestría durante el desarrollo del proyecto, en tanto que se estima que un tercer integrante, quien inició el cursado de la Maestría en Sistemas de Información el pasado año, comience la elaboración de su tesis sobre el final del desarrollo del proyecto.

Referencias

- Andresen, Lasse: Open source software is more secure than you think. SC Magazine, 2013.
- Berg, Ryan: Comment: Myths and Misconceptions on Securing Opensource Software. InfoSecurity, 2013. <http://www.infosecurity->

magazine.com/view/33945/comment-myths-and-misconceptions-on-securing-opensource-software/.

- CESSI, Cámara de Empresas de Software y Servicios informáticos: Reporte semestral sobre el Sector de Software y Servicios Informáticos de la Argentina. Reporte a diciembre 2012, Septiembre 2013. <http://www.cessi.org.ar/opssi/>.
- Cooper, Pete y Amir Nettler: Free and open source software business and sustainability models. OSS Watch, 2014. <http://oss-watch.ac.uk/resources/businessandsustainability>.
- da Rosa, Fernando y Federico Heinz: Guía práctica sobre software libre: su selección y aplicación local en América Latina y el Caribe. UNESCO, 2007.
- Ferris, Michael: Open source code and business models: More than just a license. OpenSource.com, 2014. <http://opensource.com/business/13/5/open-source-your-code>.
- Foundation, The Document: LibreOffice Certification Program. <https://www.documentfoundation.org/certification/>
- Foundation, Free Software: Free Software Foundation Certification Program. <http://www.fsf.org/licensing/compliance/ab.html>
- Group, The BSD Certification: The Standard for BSD Certification. <http://www.bsdcertification.org/>.
- Kantis, H., P. Angelelli y V. M. Koenig: Desarrollo emprendedor: América Latina y la experiencia internacional. Banco Interamericano de Desarrollo, 2004.
- Kantis, Hugo y Pablo Angelelli: Las Empresas basadas en el Conocimiento en América Latina: Factores Clave que influyen en su Surgimiento y Desarrollo.

2006.

<http://www.littec.ungs.edu.ar/pdfespa%F1ol/DT%2009-2005%20Kantis-Angelelli.pdf>.

- Kavanagh, P.: Open Source Software: Implementation and Management. Software development. Elsevier Science, 2004.
- Krishnamurthy, Sandeep: An Analysis of Open Source Business Models. En Eds. Perspectives on Free and Open Source Software. The MIT Press, 2005.
- Melo, Alberto: Los sistemas de innovación en América Latina y el Caribe. Research Department Publications 4284, Inter-American Development Bank, Research Department, 2001.
- Presidencia de la Nación Argentina, Ministerio de Ciencia, Tecnología e Innovación Productiva: Libro blanco de la Prospectiva TIC: Proyecto 2020. Artes Gráficas Papiros S.A.C.I., 2009.
- Presidencia de la Nación Argentina, Ministerio de la Industria: Plan estratégico industrial Argentina 2020 - Capítulo 12: Software y Electrónicos. Plan de Gobierno, Ministerio de la Industria - Argentina, Septiembre 2012. <http://www.industria.gob.ar/plan-estrategico-industrial-2020/>.
- Stallman, Richard M.: Free Software, Free Society: Selected Essays. GNU Press, Boston, Massachusetts, 2002.
- The Linux Foundation: Linux Adoption Trends: A Survey of Enterprise End Users. Linux Foundation Report, 2010. <http://www.linuxfoundation.org/publications/linux-foundation/linux-adoption-trends-end-user-survey>.
- Wasserman, Tony: Building a Business on Open Source Software. Carnegie Mellon University, 2009. http://repository.cmu.edu/silicon_valley/38/

Evaluaciones de Accesibilidad y Usabilidad en la WWW: Propuestas para Mejorar la Experiencia del Usuario

**Adriana Martín, Gabriela Gaetán, Viviana Saldaño, Gabriela Miranda,
Hernán Sosa, Analía, Pires, Exequiel Nichele**

Grupo de Investigación y Formación en Ingeniería de Software (GIFIS)

Instituto de Tecnología Aplicada (ITA), UNPA-UACO

Acceso Norte Ruta 3, (9011) Caleta Olivia, Santa Cruz, Argentina

e-mails: {amartin / ggaetan / vivianas / gmiranda / apires/ emolina}@uaco.unpa.edu.ar;
hassio_09@hotmail.com; eenichele@gmail.com}

Resumen

La Web se ha convertido en el medio preferido por las personas para la interacción social y de negocio. Para derribar los diferentes tipos de barreras que obstaculizan el uso de la Web y propiciar el acceso y la interacción de los diferentes perfiles de usuarios, se necesitan enfoques proactivos en el desarrollo y mejora de los productos.

El “*Grupo de Investigación y Formación en Ingeniería de Software (GIFIS)*” de UNPA-UACO está trabajando fuertemente desde el año 2010 con Proyectos de Investigación (PI) destinados a atender esta problemática nada trivial, que implica atender a las necesidades de la diversidad humana, y ha focalizado los esfuerzos de investigación en proponer soluciones que aporten a mejorar la experiencia del usuario (UX) con la Web.

Palabras clave: *Perfiles de Usuarios de Interés; Diseño y Evaluación de UI; Interfaces Móviles; Accesibilidad Web; Usabilidad Web; Web UX.*

1. Contexto

El Proyecto de Investigación (PI) N° 29/B167, Período: 2014-2016, Denominado: “*Identificación, Desarrollo y Uso de Soluciones Web Centradas en el Usuario*”, esta dirigido por la Dra. Adriana Martín y codirigido por la Mg. Gabriela Gaetán, del Instituto de Tecnología Aplicada (ITA), Universidad Nacional de la Patagonia Austral (UNPA), Unidad Académica Caleta Olivia

(UACO). Este PI se plasmó como una continuación del PI N° 29/B144, Período: 2012-2014, Denominado: “*Diseño y Evaluación de Portales Web.*” Los resultados alcanzados y la valiosa experiencia recopilada por los miembros de GIFIS durante los PI anteriores, constituyen una base sólida de antecedentes sobre los cuales fijamos el objetivo de investigación del PI N° 29/B167, focalizado en: “*Proponer técnicas y herramientas conceptuales y prácticas para desarrollar (nuevos productos) y/o mejorar (productos existentes), enfocando los esfuerzos de investigación en aquellas características que propician el acceso y la interacción del usuario con la Web.*”

Desde el comienzo de su actividad de investigación y formación, GIFIS se ha relacionado con laboratorios y grupos de investigación de otras universidades nacionales y privadas, como así también ha propiciado el intercambio con el medio. Por ejemplo, en lo que se refiere a la interacción con el medio, es importante destacar el vínculo de investigación y capacitación que mantiene con la Agencia PAMI Caleta Olivia; vínculo para el cual la Sra. Laura Pizarro es la agente PAMI local designada para la interacción GIFIS-PAMI. Así mismo, GIFIS ha asignando investigadores al dictado de los cursos de informática 2014-2015 del Programa Universidad para Adultos

Mayores, Denominado: UPAMI¹, ámbito que le permite a GIFIS llevar a cabo sus pruebas empíricas de investigación con los cursantes: nuestros abuelos de Caleta Olivia.

2. Introducción

Mejorar la experiencia del usuario en la Web (Web User eXperience), no es una tarea nada trivial, ya que implica considerar a los usuarios y sus diferentes características y capacidades. Comprometidos con esta problemática, se pueden citar organizaciones de reconocida trayectoria internacional [2][12][19][24], valiosos estándares y guías [20][24], como así también, diferentes áreas de conocimiento, tales como la Usabilidad y la Accesibilidad Web, la Interacción Hombre-Computadora (HCI) y el Diseño Centrado en el Usuario (UCD). Así, desde estos campos de trabajo, los esfuerzos de los GI y desarrollo contribuyen a mejorar las propiedades de calidad en las interfaces de usuario (UI) Web, por ejemplo, a través de la evaluación/ reparación [1][3][6][7][8][9][15][21][22][23], el diseño/ re-diseño/ adaptabilidad [4][5][10][13][14][29], entre otros.

Sin embargo, todos los esfuerzos invertidos aún resultan insuficientes para proveer soluciones que de forma efectiva y práctica, permitan manejar el abanico de perfiles de usuario Web que propone la diversidad humana. Además, las soluciones requeridas deben ser: (i) prácticas, (ii) factibles de ser transferidas y, (iii) efectivas, para que una vez implementadas, propicien el acceso y la interacción de los perfiles de usuario Web de interés.

3. Líneas de Investigación y Desarrollo

El perfil de nuestro “Grupo de Investigación y Formación en Ingeniería de Software (GIFIS)” se ha definido en base a las

actividades vinculadas con la Investigación y la Transferencia en las áreas de Ingeniería de Software e Ingeniería Web, y en particular con la temática “Web User eXperience”, muy difundida en los últimos años.

El Plan de Actividades para la ejecución del PI N° 29/B167 fue diseñado considerando las siguientes Líneas de Investigación (LI):

- LI.1 Re-diseño de sitios Web, aplicando técnicas y/o herramientas que permitan derribar barreras de Accesibilidad que enfrentan los usuarios con discapacidad visual.
- LI.2 Usabilidad y Accesibilidad para sitios Web móviles, considerando al usuario como el centro del diseño y desarrollo.
- LI.3 Soluciones para mejorar del acceso y la interacción del usuario adulto mayor con aplicaciones Web de correo electrónico, navegadores y redes sociales.
- LI.4 Técnicas para la incorporación de atributos de Accesibilidad en el desarrollo de productos Web aplicando métodos ágiles.
- LI.5 Enriquecer los lenguajes aumentativos y alternativos con propiedades de Accesibilidad y Usabilidad para mejorar la interacción de los usuarios con discapacidad que utilizan sistemas de comunicación no-verbales.

4. Resultados y Objetivos

El objetivo direccional del PI N° 29/B167 es: “Desarrollar Soluciones que propicien el Acceso y la Interacción de los Usuarios Web.” Para ello, en el marco de este PI, GIFIS se ha focalizado en desarrollar soluciones efectivas y prácticas que consideren e integren los resultados ya alcanzados en diferentes áreas de conocimiento, tales como: la Usabilidad, la Accesibilidad, los factores que contribuyen a mejorar la experiencia del usuario con la Web (“Web User eXperience”), la Interacción Hombre-Computadora (HCI) y el Diseño/Re-Diseño de Interfaces.

En la ejecución del Plan de Actividades se han considerando grupos críticos de referencia (perfiles de usuario) y los siguientes Objetivos Operacionales (OO):

¹Convenio UPAMI, UNPA-PAMI,
<http://www.uaco.unpa.edu.ar/uaco/index.php?option=com_content&view=article&id=586:renuevan-cupos-para-jubilados-interesados-en-los-talleres-upami-que-dictan-en-la-unpa&catid=33:noticias&Itemid=36>

- OO.1 Definir los requerimientos del grupo crítico de referencia: usuarios con diferentes características y/o tipos de discapacidades permanentes o temporales.
- OO.2 Identificar una solución posible para propiciar el acceso y la interacción del grupo crítico de referencia.
- OO.3 Desarrollar y validar la solución propuesta.

Con respecto a la LI.1 se ha realizado un estudio que identifica las barreras de accesibilidad para usuarios con discapacidad visual en un portal universitario y propone el rediseño del sitio incorporando patrones de accesibilidad Web. El enfoque se ha implementado en un caso de estudio real: el Portal Web de la Universidad Nacional de la Patagonia Austral (UNPA). Además, se ha realizado una evaluación de accesibilidad para comprobar la efectividad del sitio rediseñado aplicando el mismo proceso de evaluación que al sitio original. Los resultados que se han alcanzado muestran que los patrones aplicados sirven para mejorar la accesibilidad Web de personas no videntes en el portal estudiado. Los detalles de este trabajo se han reflejado en [16][17]. También, dentro de la misma LI.1 y a los efectos de dar continuidad y soporte a investigaciones previas, se ha desarrollado una experiencia completa de evaluación de un sitio Web universitario que incorpora la participación activa de usuarios no videntes. La actividad de evaluación se ha llevado a cabo en la ciudad de Río Gallegos y en un entorno familiar a los usuarios no videntes participantes, denominado: Núcleo de Acceso al Conocimiento (NAC)². Este es un espacio destinado a la realización de actividades de capacitación que dispone de conectividad inalámbrica, lo que ha posibilitado que cada participante realice las tareas de evaluación requeridas desde su propia *notebook*. La evaluación se ha desarrollado con la participación de 5(cinco) usuarios no videntes, los cuales ejecutaron 4(cuatro) tareas relacionadas con la búsqueda

de información en dicho sitio Web. Los datos recolectados y el registro de comentarios resultantes de las sesiones de Observación de Usuarios durante la evaluación del sitio Web bajo análisis, nos han permitido: (i) Determinar que el sitio tiene un nivel de Usabilidad bajo: 37,5 según la escala SUS³; (ii) Identificar nuevas barreras en el acceso e interacción de los usuarios no videntes, con respecto al conjunto de barreras de accesibilidad identificadas en estudios previos realizados por nuestro grupo de investigación; e (iii) Identificar problemas relacionados con el diseño de la Arquitectura de Información del sitio. Los hallazgos se han plasmado en las siguientes contribuciones [18][27].

En la LI.2, nos hemos concentrado en definir el enfoque aplicado en [11] al desarrollo de un prototipo de sitio Web móvil para UNPA-UACO. El enfoque propuesto se ha centrado en el usuario incorporando propiedades de Usabilidad y Accesibilidad para obtener una IU que pueda ser desplegada y utilizada correctamente en los dispositivos móviles actuales. El resultado de este trabajo se ha plasmado en la siguiente contribución [26].

Por otra parte, en la LI.3 se ha trabajado en el Proyecto de Cultura Científica, Código CC16/15, Denominado: “*Por una Sociedad inclusiva en la que los Adultos Mayores se involucren con la Web.*” El objetivo de este proyecto ha sido difundir consejos prácticos que, con un lenguaje simple y fácil de seguir, puedan propiciar una navegación en la Web mas confiada y autónoma por parte de los adultos mayores. A tal efecto, se han diseñado en formato papel, algunas guías prácticas para ayudar a los adultos mayores a resolver los problemas que más frecuentemente encuentran al navegar por la Web. Estas guías prácticas se han distribuido en las oficinas de PAMI, donde los adultos mayores acuden cotidianamente.

² NAC Río Gallegos, Santa Cruz
<<http://www.nac.gob.ar/busca-tu-nac/interior/id/76>>

³ System Usability Scale (SUS)
<<http://www.sidar.org/recur/desdi/traduc/es/visitable/nuevos/CuestCon.htm>>

También enfocados en atender la LI.3 se ha desarrollado el Taller: “*Usos de Internet en la Vida Cotidiana*” destinado a adultos mayores. El objetivo académico de este proyecto estuvo orientado a profundizar el conocimiento y la familiaridad con herramientas de la Web (navegadores, buscadores, herramientas de comunicación, gestión de trámites, comercio electrónico). Mientras que desde el punto de vista investigativo, se ha fijado como objetivo recolectar información, mediante la observación directa de los usuarios adultos mayores, sobre dificultades de usabilidad, dificultades de accesibilidad (relacionadas con problemas de visión, audición, movilidad, dificultades cognitivas), y sobre los inconvenientes más frecuentes con los que se enfrentan los adultos mayores al utilizar la Web.

Durante 2015, GIFIS ha aportado las siguientes contribuciones científicas en los ámbitos Nacional e Internacional [18][25][26][27][28].

5. Formación de Recursos Humanos

GIFIS reúne a 7(siete) investigadores, entre docentes y alumnos de grado y postgrado UNPA, según el siguiente detalle: 2(dos) docentes de grado y postgrado de las carreras Ingeniería en Sistemas (IS-UNPA) y Maestría en Informática y Sistemas (MIS-UNPA); 2(dos) docentes de grado de IS-UNPA; 1(un) alumno de postgrado de MIS-UNPA y, 2(dos) alumnos de grado IS-UNPA.

La directora y codirectora del PI N°: 29/B167, Dra. Adriana Martín y Mg. Gabriela Gaetán respectivamente, son Categoría III en el Programa de Incentivos Docentes. La docente-investigadora Mg. Viviana Saldaño es Categoría IV en el Programa de Incentivos Docentes.

Las 4 (cuatro) docentes integrantes de GIFIS, se han presentado a la V Convocatoria a Categorización del Programa de Incentivos a Docentes Investigadores.

Durante 2015, fueron evaluadas favorablemente por revisión de pares 2(dos) becas de investigación de grado ejecutadas durante 2014, de los alumnos de IS-UNPA Hernán Sosa y Diego Vilte, denominadas “*Derribando Barreras en la Interacción de los Usuarios Web con Discapacidad*” y “*Derribando Barreras en la Interacción de los Usuarios Web Adultos Mayores*”, respectivamente.

Durante 2015, la Ing. Gabriela Miranda, alumna de postgrado de MIS-UNPA, entregó a evaluación su Proyecto de Tesis denominado: “*Un Enfoque de Desarrollo Basado en la Experiencia del Usuario (UX) para Favorecer la Accesibilidad Web*”. Esta tesis está siendo dirigida por Dra. Adriana Martín y codirigida por la Mg. Gabriela Gaetán.

Actualmente, se está dirigiendo y codirigiendo 1 (una) tesis de postgrado MIS-UNPA y también, se está codirigiendo 1 (una) tesis UNICEN de la carrera Doctorado en Ciencias de la Computación.

6. Agradecimientos

A la UNPA por el soporte al Proyecto de Investigación (PI) N°: 29/B167, Período: 2014-2016, Denominado: “*Identificación, Desarrollo y Uso de Soluciones Web Centradas en el Usuario*”

7. Referencias

- [1] Brajnik, G. Barrier Walkthrough: Heuristic Evaluation Guided by Accessibility Barriers<<http://users.dimi.uniud.it/~giorgio.brajnik/projects/bw/bw.html>>
- [2] Fundación SIDAR: Acceso Universal <<http://www.sidar.org/recur/direc/legis/espa.php>>
- [3] Martín, A., Gaetán, G., Saldaño, V., Miranda, G., Vilte, D., Sosa, H., Molina, S. Identificación, Desarrollo y Uso de Soluciones Web Centradas en el Usuario. WICC 2014.
- [4] Martín, A., Miranda, G., Saldaño, V., Gaetán, G. AO-WAD: A Generalized Approach for Accessible Design within the Development of Web-based Systems. The 7th International Conference on Software Engineering Advances, ICSEA 2012, Nov 18-23, Portugal. ISBN:978-1-61208-230-1.

- [5] Martín, A., Miranda, G., Saldaño, V., Gaetan, G. AO-WAD: A Proposal for Accessible Design within Web Engineering Approaches. CACIC 2012.
- [6] Martín, A., Gaetán, G., Saldaño, V., Miranda, G., Pires, A., Vilte, D., Sosa, H., Molina, S. Identificación, Desarrollo y Uso de Soluciones Web Centradas en el Usuario. 3er. Encuentro de Invest. UNPA-UART 2014.
- [7] Miranda, G., Martín, A., Mazalú, R., Gaetán, G. Y Saldaño, V. Agentes Inteligentes para propiciar la Accesibilidad Web. CACIC 2013.
- [8] Miranda, G., Martín, A., Gaetán, G. Mejora de la Accesibilidad Web mediante el Uso de Agentes Inteligentes. Una Revisión del Estado del Arte. ICT-UNPA 2013, Vol.5, N°2, P133-160. ISSN:1852-4516.
- [9] Miranda, M. G., Martín, A., ; Saldaño, V., Gaetán, G. Usabilidad y Accesibilidad en la Redes Sociales. ICT-UNPA 2014, Vol.6, N°2, P118-140. ISSN:1852-4516.
- [10] Moreno, L., Martínez, P., Ruiz, B. A MDD Approach for Modeling Web Accessibility. IWOST 2008.
- [11] Moyano, A. Interface Móvil para el Sitio Web de la UACO: Un Prototipo basado en la Experiencia del Usuario. Tesis Ingeniería en Sistemas, UNPA-UACO, 2014.
- [12] Nielsen Norman Group. Evidence-Based User Experience Research, Training, and Consulting<<http://www.nngroup.com/articles/definition-user-experience/>>
- [13] Plessers P., Casteleyn S. , Yesilada Y. , De Troyer O. , Stevens R. , Harper S., Goble C. Accessibility: A Web Engineering Approach. WWW 2005.
- [14] Saldaño, V., Martín, A., Gaetán, G. y Vilte, D. Web Accessibility for Older Users: A Southern Argentinean View. 8th International Conference on Software Engineering Advances ICSEA 2013, Oct. 27-Nov.1, Italy. ISBN:978-1-61208-304-9.
- [15] Saldaño, V., Martín, A., Gaetán, G., Vilte, D. Focusing on Older Web Users: An Experience in Patagonia Argentina. International Journal on Advances in Internet Technology. IARIA 2014, Vol.7, N°3&4, P172-182. ISSN:1942-2652.
- [16] Sosa, H. Rediseño de un Sitio Web Universitario usando Patrones de Accesibilidad Web. CoNaIISI 2014.
- [17] Sosa, H., Gaetán, G., Martín, A., Saldaño, V. Propuesta Basada en Patrones para Mejorar la Accesibilidad de Usuarios Ciegos en un Sitio Web Universitario. 3er. Encuentro de Invest. UNPA-UART 2014.
- [18] Gaetán, G., Martín, A., Saldaño, V., Sosa, H. Identifying Accessibility Barriers of a University Website: An Evaluation from the Perspective of Blind Users. Sixth International Conference on Advances in New Technologies, Interactive Interfaces and Communicability ADNTIIC 2015. ISBN 978.88.96.471.43.2
- [19] The Paciello Group: Your Accessibility Partner<<http://www.paciellogroup.com/>>
- [20] US Section 508: Opening Doors to IT. <<http://www.section508.gov/?fuseAction=stdsdoc> >
- [21] Vilte, D., Saldaño, V., Martín, A., Gaetán, G. Evaluación del Uso de Redes Sociales en la Tercera Edad. CoNaIISI 2013.
- [22] Vilte, D., Martín, A.; Gaetán, G., Saldaño V. Favoreciendo el Acceso a la Web: Experiencias con Usuarios Mayores de la Región Patagónica. ICT-UNPA 2014, Vol:6, N°2, P1-29. ISSN:1852-4516.
- [23] Vilte, D., Saldaño, V., Martín, A., Gaetán, G. Conociendo las Barreras de Interacción Web en la Tercera Edad. 3er. Encuentro de Invest. UNPA-UART 2014.
- [24] W3C-WAI Project <http://www.w3.org/WAI/>
- [25] Martín, A., Gaetán, G., Saldaño, V., Miranda, G., Sosa, H., Vilte, D., Pires, A., Molina, S. Hacia una WWW a la medida de las Personas: Identificación, Desarrollo y Uso de Soluciones Web Centradas en el Usuario. WICC 2015.
- [26] Moyano, A., Gaetán, G., Martín, A. Interfaz Móvil para el Sitio Web de la UACO. Un Prototipo centrado en el Usuario. ICT-UNPA-129-2015.
- [27] Sosa, H., Gaetán, G., Martín, A. Rediseño de un Portal Web Universitario aplicando Patrones de Accesibilidad. ICT-UNPA 2015, Vol.7, N°2, P137-166. ISSN:1852-4516.
- [28] Vilte, D., Saldaño, V., Gaetán, G., Martín, A. Identificando Barreras en la Interacción con Facebook: Una Experiencia con Adultos Mayores de la Patagonia Austral. ICT-UNPA 2015, Vol.7, N°2, P249-268. ISSN:1852-4516.
- [29] Martín, A. Engineering Accesible Web Applications: An Aspect-Oriented Approach. Libro. EDULP UNLP, 2014, Págs.206, ISBN:978-987-1985-28-9

Metodologías para el desarrollo de software en PYMES

Marcelo A. Colombani, Martín M. Pérez, Cristian D. Pacífico

Lenguajes de Programación e Ingeniería en Software / Facultad de Ciencias de la Administración / Universidad Nacional de Entre Ríos (UNER)

Monseñor Tavella 1424, (0345) 4231400

marcol@fcad.uner.edu.ar, marper@fcad.uner.edu.ar, cripac@fcad.uner.edu.ar

Resumen

Uno de los principales objetivos de este proyecto de investigación en el cual está inmerso el presente trabajo, es especificar modelos y guías de certificación para productos de Software Libre y procesos de ciclo de vida que involucren plataformas open source.

Para alcanzar los objetivos propuestos se abordaron los diferentes modelos de proceso de software, desarrollando sus características, bondades y deficiencias, con el fin de detectar qué metodología se encuadra o adapta mejor al escenario de las Pequeñas y Medianas Empresas productoras de software o empresas que en su estructura albergan un equipo de desarrollo no demasiado numeroso.

Se realizará un relevamiento en PYMES en la Región de Salto Grande, que tengan grupos de trabajos dedicados al proceso de ciclo de vida de software.

Palabras clave: PYMES, proceso de desarrollo de software, modelos de proceso de software, modelos y guías de certificación.

Contexto

El trabajo forma parte de la línea de investigación del Proyecto de Investigación y Desarrollo (PID)

denominado “*Modelos de Certificación para los procesos de Ingeniería de Software en productos desarrollados con Lenguajes de Programación Open Source: relevamiento y aplicación en PYMES de la zona de influencia de la UNER Concordia*” presentado en la Fac. Cs Administración de la UNER.

El PID busca establecer las bases del ámbito científico-académico para el desarrollo y redacción de guías y normas para la validación y certificación de productos y procesos de Software Libre y Tecnologías Abiertas.

Dentro del marco de la presente línea de investigación se está desarrollando una tesis de maestría, en la carrera *Maestría en Sistemas de Información* dictada en la Fac. Cs Administración, UNER en el área de Ingeniería del Software.

Introducción

El software de computadoras se ha convertido en un producto fundamental en todas las empresas y organizaciones, es un factor clave que diferencia a las que lo posean. Actualmente existen sistemas en casi todo tipo de empresa y ellas poseen una gran variedad de los mismos, como ser: transporte, gobierno, medicina, agro, telecomunicaciones, herramientas de oficina, entre tantos, y la lista se hace interminable [Som15].

El impacto del software en nuestra sociedad y en la cultura continúa siendo profundo. Al mismo tiempo que crece su importancia, la comunidad del software trata continuamente de desarrollar tecnologías que hagan más sencilla, rápida y menos costosa la construcción de programas de computadora de alta calidad [PM15].

La utilización masiva de las computadoras y por ende de los sistemas en los negocios ha generado una cultura de “sistematización de los procesos”. Sumado a esto, las herramientas de desarrollo han permitido la generación de software de una manera mucho más fácil, rápida y menos costosa, lo que ha sustentado la cultura de sistematización. Así se logró insertar sistemas software en empresas o negocios pequeños, que de otra manera no hubiesen podido encarar un proyecto de este tipo.

A raíz de esto cada vez se requieren más y mejores sistemas y en tiempos considerablemente menores, pero la construcción de software no es tarea fácil por varios motivos. Normalmente se requiere que el software cubra todas las necesidades de la empresa, que se desarrolle en un tiempo limitado y, además, existen innumerables metodologías de desarrollo de software que sustentan este proceso de construcción. Existen diferentes propuestas tradicionales [PM15], [Som15] que se direccionan especialmente en el control del proceso, definiendo rigurosamente las actividades a desarrollar, los resultados a producir, las notaciones y las herramientas a utilizar. Como alternativa para mejorar el desarrollo se pueden incluir más actividades [Som15], más diagramas [Som15], pero esto sólo puede llevar a un

desarrollo más complejo y a sobrecargar al equipo de trabajo.

Una alternativa a este proceso de desarrollo se basa en centrarse en otros factores, como son el humano o el producto software a construir. Esta idea se conoce como Metodologías Ágiles [Bee09], que dan mayor valor al individuo, mayor participación al cliente, y tienen como objetivo realizar entregas incrementales del software.

Estos métodos son utilizados en proyectos en los cuales los requerimientos o expectativas del clientes/usuario son cambiantes o difusos. También suele exigirse un tiempo de entrega reducido con un producto que respete estándares de calidad [Som15].

En experiencias en desarrollo de software surge una problemática recurrente: la falta de adecuación de los requerimientos establecidos de un software a desarrollar con las expectativas que el cliente/usuario tiene del mismo [ISO08], [Som15]. Esto tiene su origen en que los métodos tradicionales que exigen que los requerimientos sean establecidos rigurosamente; en contraposición con las expectativas del cliente/usuario las cuales son dinámicas a lo largo del proyecto. De hecho, tales expectativas se modifican a medida que se visualizan las potencialidades del software desarrollado en las entregas parciales del mismo; entregas que por otro lado, son necesarias para verificar tempranamente el producto de software y renovar el compromiso hacia el proyecto [PM15], [Bee09], [Wel09]. A medida que el proyecto se desarrolla, es posible que otros interesados (*stakeholders*) intervengan en el mismo, y vean potencialidades en el dominio de aplicación no contempladas en los requerimientos. En este escenario,

es habitual que la empresa cliente vea que algunas expectativas no se cumplen con la versión final del producto; por más que éste cumpla plenamente los requerimientos. Por ende, resulta necesario para la aceptación del producto (y la impresión positiva del equipo de desarrollo), reflejar estas expectativas no cubiertas en los requerimientos de alguna manera, para que se materialicen en la versión final del producto.

Esta problemática ha llevado a la necesidad de utilizar diferentes modelos de proceso de software adaptados y combinados de diferente manera para lograr un proceso óptimo sin perder de vista uno de los objetivos que es lograr un software de calidad en todos los aspectos, y confiable en todas sus dimensiones tanto en disponibilidad, fiabilidad, seguridad y protección.

Modelos de proceso

Los modelos son simplificaciones de los procesos de software; por lo tanto, un modelo de procesos del software es una simplificación o abstracción de un proceso real [Som15]. Se puede definir a un modelo de procesos del software como una representación abstracta de alto nivel de un proceso de software [Som15].

Revisando los enfoques propuestos ([Som15], [PM15], [Wells09], [JBR99]) es posible identificar una variedad de modelos de desarrollo de software, como así también modelos similares mencionados de diferente manera, pero realizando un análisis minucioso de su conformación y manera de organizar el proceso, éstos realmente reflejan el mismo funcionamiento.

Por lo anteriormente expuesto, en este trabajo se enumeran y describen los

diferentes modelos que servirán de guía para clarificar qué modelo de desarrollo puede ajustarse mejor a las Pequeñas y Medianas Empresas de nuestra zona (Concordia, Entre Ríos, Argentina), las cuales no están ajenas al problema de los requerimientos confusos, incompletos y cambiantes.

Cada modelo es una descripción de un proceso software que se presenta desde una perspectiva particular [Som15], describiendo una sucesión de fases y una relación entre ellas. Según las fases y el modo en que se produzca esa relación, tenemos diferentes modelos de proceso. Es por ello que un modelo es más adecuado que otro para desarrollar un proyecto dependiendo de un conjunto de características de éste [Som15], [PM15]. Análogamente, los artefactos entregables de cada etapa varían conforme su tasa de cambio y validación; en el caso que se acepte la dinámica de requerimientos como una suposición presente en el modelo, estos cambios deben documentarse y trazarse adecuadamente, para justificar posibles desviaciones no contemplada en el proyecto.

Otra faceta que no debe dejarse de lado, es el aseguramiento de la calidad en el proceso y producto de software; en particular si es un requisito para certificar o validar conforme a estándares internacionales de calidad [Flo13]. Como equipo (empresa) de desarrollo resulta adecuado instaurar buenas prácticas en el proceso de desarrollo conforme a guías de certificación aceptadas (ISO/IEC 9001-90003 [ISO15, ISO14], ISO 15504 SPICE [ISO04], ISO/IEC 15288 [ISO08], ISO/IEC/IEEE 12207 y complementarias [ISO08b, ISO11c]). Como empresa usuaria o productora de software sería importante verificar la calidad del producto conforme a esquemas de

certificaciones (ISO/IEC 25000 [ISO07, ISO11, ISO11b]), permitiendo como efecto deseado tener ventajas competitivas y comerciales.

Resultados Parciales

En base a la investigación realizada se puede asegurar que los cambios son inevitables en todos los proyectos de software, sean éstos pequeños, medianos o grandes. Los requerimientos del sistema cambian cuando el negocio donde está inserto el sistema responde a las presiones externas.

Las prioridades de gestión cambian y, cuando se dispone de nuevas tecnologías, cambian los diseños y la implementación. Esto significa que el proceso del software no es un proceso único; más bien, las actividades del proceso se repiten regularmente conforme el sistema se rehace en respuesta a peticiones de cambios.

Como surge del estudio de las diferentes metodologías, el problema de cada una de ellas siempre gira alrededor de la falta de requerimientos, la gestión de la documentación y la falta de comprensión del cliente con respecto a los requerimientos. Todos estos factores llevan a requerimientos confusos, incompletos y cambiantes.

Además de la metodología que se utilice, será importante definir una buena estructura de documentación, a fin de poder reflejar los requerimientos que dieron origen al sistema, los diferentes componentes del sistema software, sus relaciones y sus efectos en las modificaciones del mismo.

Es importante también detallar una guía de buenas prácticas que aseguren la

calidad en el proceso de desarrollo y en el producto software final, siguiendo los esquemas de normas internacionales. En particular se debe puntualizar detalladamente cómo los cambios en los requerimientos impactan en el proceso y en la calidad del producto, y de qué forma medir y registrar dichos efectos, para lograr mejorar continuamente el modelo de ciclo de vida.

Líneas de Investigación, Desarrollo e Innovación

Como trabajo futuro será importante continuar el estudio de los Modelos de Proceso de Software y describir pasos metodológicos para llevar adelante un proyecto de software exitoso utilizando alguna combinación de diferentes metodologías tradicionales y ágiles, ya que cada una posee ventajas y desventajas, logrando así un enfoque híbrido en el cual los métodos ágiles incorporen algunas técnicas de planificación y documentación pero sin perder de vista las iteraciones y los principios del modelo ágil.

Resultados y Objetivos

Este proyecto de investigación se encuentra en su fase de iniciación y, entre los objetivos buscados podemos remarcar la identificación de diversas metodologías de desarrollo de software que pueda ser utilizada en empresas de nuestra zona (Concordia, Entre Ríos, Argentina), que cuenten con un grupo de desarrolladores reducido o en muchos casos unipersonal, y consideren llevar adelante un proceso de desarrollo organizado y de calidad. Por otro lado, se pretende redactar guías y manuales para la aplicación de tales métodos y que permitan guiar en la adopción de normas y certificación de

estándares del proceso de desarrollo y del producto final de software.

Formación de Recursos Humanos

Como parte de este proyecto de investigación se espera completar una tesis de maestría en la Maestría en Sistemas de Información. Los integrantes están dirigiendo tesis de grado de la Facultad Regional de Concordia, Universidad Tecnológica Nacional y dirigiendo una Tesis de Maestría en la Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos.

Referencias

[Bee09] Beedle, M., “Principles behind the Agile Manifesto”. [online]. <http://www.agilemanifesto.org/principles.html>

[Flo13] Florian Schneider, Brian Berenbach, “A Literature Survey on International Standards for Systems Requirements Engineering”, *Procedia Computer Science*, Volume 16, 2013, Pages 796-805.

[ISO04] ISO, “ISO/IEC 15504 Information technology - Process assessment” ISO/IEC, Nov, 2004.

[ISO07] ISO, “ISO/IEC 25030 - Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Quality requirements,” Jun. 2007.

[ISO08] ISO, “ISO/IEC 15288:2008 - Systems and software engineering - System life cycle processes.” Mar. 2008.

[ISO08b] ISO, “ISO/IEC/IEEE 12207:08 Systems and software engineering - Software life cycle processes,” Mar. 2008.

[ISO10] ISO, “ISO/IEC/IEEE 24765:10 - Systems and software engineering - Vocabulary”, Dec. 2010.

[ISO11] ISO, “ISO/IEC 25040 – Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Evaluation process.” Feb. 2011.

[ISO11b] ISO, “ISO/IEC 25010 –Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models,” Mar. 2011.

[ISO11c] ISO, “ISO/IEC/IEEE 29148: 2011 - Systems and software engineering - Life cycle processes - Requirements engineering,” Nov. 2011.

[ISO14] ISO, “ISO/IEC 90003:2014 - Software engineering - Guidelines for the application of ISO 9001:2008 to computer software” Dec, 2014.

[ISO15] ISO, “ISO/IEC 9001:2015 - Quality management systems - Requirements” Sep, 2015.

[JBR99] I. Jacobson, G. Booch, and J. Rumbaugh. “El proceso unificado de desarrollo de software”. Addison-Wesley, primera edición, 1999.

[Pal07] Juan Palacio. “Flexibilidad con scrum”. Safe Creative, 2007.

[PM14] R.S. Pressman, B.Maxim. “Software Engineering: A Practitioner's Approach”. Mc Graw Hill, 7ma Ed, 2014.

[Som15] I. Sommerville. “Software Engineering” (10th Ed). Pearson Addison Wesley, 2015.

[Wel09] D. Wells. “Extreme programming: A gentle introduction”. Tech. Report, 2009. [online] <http://www.extremeprogramming.org>

Elicitación de requerimientos con *Grounded Theory*

Alejandro Oliveros, Ana Piccin, Diana Cicinelli

Facultad de Ingeniería y Tecnología Informática/ Universidad de Belgrano

Av. F. Lacroze 1947, CABA

aoliveros@ub.edu.ar, {ana.piccin; diana.cicinelli}@comunidad.ub.edu.ar

Resumen

La *Grounded Theory* es una metodología utilizada en las ciencias sociales. Este trabajo propone determinar si esta metodología resulta efectiva para la práctica de la elicitación de requerimientos, si el modelo que emergerá de su aplicación será válido para este propósito. Surgen de este modo dos objetivos. En primer lugar, aplicar la GT a los datos básicos de los cuales se extraerá el modelo de requerimientos. En segundo lugar, validar el modelo obtenido. El proyecto se encuentra actualmente en desarrollo y se está en vías de la construcción del modelo. Se describen los pasos seguidos y las conclusiones obtenidas hasta el momento.

Palabras clave: Elicitación de requerimientos, Lenguaje natural, *Grounded Theory*.

Contexto

Elicitación de requerimientos

Mediante la Elicitación de Requerimientos se obtiene el conocimiento relevante para producir un modelo de requerimientos de un determinado dominio [1]. Parte significativa de este conocimiento se encuentra en Lenguaje Natural (LN), la mayoría de los documentos disponibles para el análisis de requerimientos son

provistos por el usuario o resultan de entrevistas [2]. Los enfoques de elicitación con LN se pueden clasificar en aquellos que interactúan directamente con el usuario y los que lo hacen con documentos [1]. La interacción con el usuario genera nuevamente documentos en LN en los que se registra esa interacción. En documentos de los sistemas heredados, en general formulados en LN [3], como manual del usuario o especificaciones de requerimientos, hay información relevante para la construcción de un nuevo sistema.

Pese a aspectos tales como la ambigüedad o la informalidad, el LN resulta esencial y de difícil reemplazo en el proceso de capturar el conocimiento necesario para producir la Especificación de Requerimientos de Software (SRS). En determinados momentos en los que aún no está consolidado el conocimiento del dominio, mantener una cierta ambigüedad puede resultar útil para avanzar en el proceso de elicitación y a los *stakeholders* puede resultarles conveniente un modo de expresión informal. El conocimiento del significado de los términos no específicos utilizados y de sus sintaxis son aspectos que favorecen la utilización del LN. En general el LN supera a los lenguajes formales o semi-formales en el alcance de su capacidad de descripción.

A pesar de que se dispone de una importante gama de técnicas de

elicitación de requerimientos y de métodos que proveen guías para su utilización [4], se necesitan técnicas que faciliten el tratamiento de descripciones inconsistentes e incompletas [4].

El proceso de elicitación de requerimientos de un sistema software utiliza una variedad de fuentes que debe procesar para producir la SRS. La elicitación produce conocimiento y ese conocimiento es el principal input del proceso que produce la SRS. No es lo mismo la información del mundo real (minutas, documentos, actas, especificaciones, etc.) que el conocimiento que se elabora en el proceso de elicitación. Establecida la diferencia cabe preguntarse: ¿dónde reside ese conocimiento?, ¿qué artefactos soportan ese conocimiento? Existen opciones como la construcción del Léxico Extendido del Lenguaje [4] a partir del Universo del Discurso, que permite elaborar Escenarios [5] o Casos de Uso [6].

Enfoques y técnicas provenientes de las ciencias sociales

La Ingeniería de Requerimientos (RE, por sus iniciales en inglés) se desarrolla en un contexto de actividad humana organizada en términos sociales. Una parte importante de sus *stakeholders* son seres humanos, de allí que se haya apelado a fundamentos y técnicas provenientes de disciplinas como la psicología cognitiva, antropología, sociología y lingüística [7]. La RE ha incorporado una variedad de enfoques provenientes del campo de las ciencias sociales, en especial en el terreno de la elicitación. Una pequeña lista informal incluye: *Activity Theory*, *Brainstorming*, *Etnografía*, *Focus group*, *Lingüística*.

Esta lista, no exhaustiva, permite observar que los recursos de las ciencias

sociales que se han importado al campo de RE actúan en el terreno de la práctica específica del usuario (*Activity Theory*), la interacción directa con el usuario (*Brain Storming* y *Focus Group*), el entendimiento del usuario y del contexto en el cual se desempeña (*Etnografía*), y la comprensión de los documentos del proceso de desarrollo (*Lingüística*).

Introducción

Grounded Theory (GT) se desarrolló en el campo de la sociología con el fin de descubrir teorías a partir de un análisis particular de los datos [8]. Ha sido utilizada como recurso de investigación en el campo de la ingeniería de software en general y de la RE en particular. En [9] se detalla una lista de este tipo de investigaciones en ambos campos, al que denominan, de *uso metodológico*.

Se dispone de un reducido grupo de trabajos iniciales que han utilizado GT en la RE, no como un recurso metodológico de investigación sino como una herramienta de desarrollo de software [9]: integración de etnometodología y GT (para obtención de datos y de análisis respectivamente); elicitación y análisis de requerimientos para la construcción del Modelo del Dominio; obtención de Clases, Diagramas de Flujo de Datos y Diagramas Entidad Relación; elaboración de Casos de Uso; producción de escenarios; extracción de los requerimientos funcionales y no funcionales de la Arquitectura de la Empresa y de la descripción del sistema.

En términos generales la GT provee un conjunto de herramientas que ayudan a la modelización y descripción del sistema en construcción.

Líneas de Investigación, Desarrollo e Innovación

El trabajo se enfoca en la utilización de GT como una herramienta de desarrollo de software, en el ámbito de RE. Dentro de este contexto se desarrollan tareas de apropiación de la metodología, para lo que se tienen en cuenta dos ejes que se desarrollan simultáneamente: el metodológico y el específico del dominio.

El primero incluye la transposición metodológica del dominio de las ciencias sociales al de la elicitación de requerimientos. El segundo guía la construcción del modelo de requerimientos desde el dominio de la Ingeniería de Software.

La inclusión de los *stakeholders*, como objeto de estudio de la Ingeniería de Software, ha alentado a utilizar metodologías de investigación de orden cualitativo-interpretativo, propias del dominio de las ciencias sociales. La aplicación de la GT a otros dominios, en este caso el de la elicitación de requerimientos, exige un estudio profundo que asegure la validez de los resultados.

La aplicación de la GT tiene en cuenta las siguientes instancias: Recolección de datos, Codificación, Muestreo, Comentarios (memos), Clasificación y Redacción. El orden de las instancias no es estricto, algunas de ellas pueden desarrollarse simultáneamente. Se debe aplicar la “comparación constante”, la que obliga a la revisión permanente del trabajo realizado. [15]

La GT, aplicada rigurosamente, permite que emerjan hipótesis de los datos. Para el caso de su aplicación a la ingeniería de requerimientos esto exige, por un lado, el esfuerzo de despojarse de la inercia de los métodos de investigación tradicionales del dominio. Por el otro, y a lo largo de todo el análisis, mantener la atención en no importar hipótesis de otros entornos.

Para asegurar la validez del estudio fue necesario: 1) capacitarse el grupo de trabajo para desarrollar criterios que permitieran determinar la escuela que guiaría el estudio: ya sea la clásica elaborada por Glaser y Strauss y continuada por el primero [8] o la alternativa desarrollada por Strauss y Corbin [10], optándose por la primera; 2) leer críticamente y en profundidad la bibliografía fundamental de la GT para desarrollar las competencias necesarias para su aplicación a los fines de este estudio; 3) revisiones y contrastaciones de términos e interpretaciones a lo largo de toda la aplicación de la metodología.

La elicitación de requerimientos se basó en el estudio del documento que constituye un contrato de auto-ahorro. La aplicación sistemática de la GT permitió que, a través de las distintas instancias de abstracción, surgieran las categorías sobre las que, finalmente, se organizará el modelo.

La pregunta de investigación de mayor alcance es: “¿es posible obtener el conocimiento necesario para producir la SRS mediante la utilización de GT?”. La hipótesis se puede formular como: “El conocimiento que se obtiene del proceso de elicitación de requerimientos puede alcanzarse aplicando GT”.

Resultados y Objetivos

Se dispone de un trabajo exhaustivo para la elicitación de los requerimientos en un dominio dado mediante la elaboración de LEL. Se trata del análisis del mismo universo: el sistema de Planes de Ahorro Previo para la Adquisición de Vehículos 0km sobre la base del documento “Solicitud de adhesión. Condiciones generales de contratación”, que es un contrato tipo para un sistema de autoahorro [11]. Sobre este se realizó un trabajo de aplicación de GT para luego

cotejar los resultados obtenidos con. A continuación se describen las actividades realizadas.

Objetivo. Mediante la aplicación de GT obtener el conocimiento suficiente del dominio para especificar los requerimientos.

Selección de las herramientas informáticas. Se utilizó una combinación de procesador de texto y hojas de cálculo. Ello permitió conocer en profundidad los requerimientos de una herramienta automatizada de soporte que se aspira a diseñar e implementar.

Organización de la información. Se utilizó una tabla con múltiples columnas (Texto original; Código; Concepto; Categoría) para relacionar las sucesivas codificaciones y anotaciones de cada fragmento de texto.

El proyecto se encuentra en desarrollo y, hasta la fecha, se han logrado los siguientes objetivos de análisis:

Recolección de datos: Se identificaron en el texto original los fragmentos a ser analizados y se organizaron en la columna “Párrafo”. En el proceso estos fragmentos fueron subdivididos y/o fusionados

Codificación: A cada fragmento se le asignó (“emergió”) un *código*, entendido como término que refleja el significado del texto. La ejecución de la *comparación constante*, propia de la metodología, provocó que los códigos fueran revisados reiteradamente para mantener su consistencia.

Conceptos. Mediante el análisis de los códigos emergieron conceptos que, comparados por igualdad o diferencias, permitieron obtener las *categorías*.

Muestreo teórico. Ocurre durante la redacción final del modelo. Garantiza el vínculo entre la teoría en construcción y los datos originales. Aún no se ha alcanzado esta etapa en el trabajo.

Comentarios. Durante el proceso de comparación se registraron los

comentarios de los investigadores que hacen a las razones y comprensión de los códigos, conceptos y categorías. Dado que el estudio está dirigido a la elicitación de requerimientos, la redacción de los comentarios se orientó a desarrollar textos y fórmulas que pudieran ser utilizados en la formulación de los requerimientos. En GT, al concluir el proceso de análisis de los documentos, estos memos son el material de trabajo de los investigadores.

Clasificación. Durante el proceso de comparación constante, a partir de los códigos, emergieron categorías. A medida que las categorías iban surgiendo, se comenzó a percibir el modelo.

Un recurso que se utiliza a lo largo de todo el proceso es la *comparación constante*. La identificación de códigos es un proceso de abstracción sobre el texto original. A medida que se progresa en la lectura del texto y se identifican nuevos códigos, surge la necesidad de rever los que ya han sido definidos. Esto lleva a la modificación de algunos términos para asegurar la lógica de abstracción a lo largo del proceso. De este modo, la relación entre código y párrafo tendrá la misma calidad a lo largo de todo el análisis y se facilitarán las necesarias iteraciones, o reconstrucciones de las interpretaciones, requeridas por la metodología. El mismo enfoque se aplica a la revisión de conceptos, categorías y memos.

Formación de Recursos Humanos

Se encuentra en desarrollo el trabajo final de carrera “Requerimientos funcionales de las aplicaciones educativas”, que se desarrollará mediante la aplicación de GT.

Equipo de trabajo. Lic. Alejandro Oliveros, director; Lic. Ana M. Piccin,

profesora investigadora; Lic. Diana Cicinelli, profesora investigadora.

Referencias

- [1] P. Loucopoulos and V. Karakostas, *Systems Requirements Engineering*. McGraw-Hill, 1995.
- [2] D. M. Berry and E. Kamsties, “Ambiguity in Requirements Specification,” in *Perspectives on Software Requirements*, Springer US, 2004, pp. 7 – 44.
- [3] I. John and J. Dörr, “Elicitation of requirements from user documentation,” in *REFSQ’03 - PRE-PROCEEDINGS. 9th International Workshop on REQUIREMENTS ENGINEERING - FOUNDATION FOR SOFTWARE QUALITY*, Klagenfurt/Velden, Austria, 2003, pp. 3 – 12.
- [4] J. Leite and A. Franco, “A strategy for conceptual model acquisition,” in *Proceedings of the First IEEE International Symposium on Requirements Engineering*, San Diego, California, 1993, pp. 243 – 246.
- [5] J. C. S. do P. Leite, G. Rossi, F. Balaguer, V. Maiorana, G. Kaplan, G. Hadad, and A. Oliveros, “Enhancing a Requirements Baseline with Scenarios,” *Requirements Engineering Journal*, no. 2, pp. 184 – 198, 1997.
- [6] L. Antonelli, J. C. S. do P. Leite, G. Rossi, and A. Oliveros, “Deriving requirements specifications from the application domain language captured by Language Extended Lexicon,” in *Proceedings of Workshop on Requirements Engineering*, Buenos Aires, Argentina, 2012.
- [7] B. Nuseibeh and S. Easterbrook, “Requirements Engineering: A Roadmap,” in *ICSE ’00 Proceedings of the Conference on The Future of Software Engineering*, Limerick, Ireland, 2000, pp. 35 – 46.
- [8] Glaser, Barney G. and A. L. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New Jersey, USA: Aldine Transaction, 2008.
- [9] D. M. Berry, M. W. Godfrey, R. Holt, C. J. Kapsner, and I. Ramos, “Requirements Specifications and Recovered Architectures as Grounded Theories,” *The Grounded Theory Review*, vol. 12, no. 1, pp. 56 – 65, Jun. 2013.
- [10] Strauss, Anselm L. and Corbin, Juliet M., *Basics of Qualitative Research : Techniques and Procedures for Developing Grounded Theory*. Sage Publications, Inc., 1998.
- [11] C. Litvak, G. Hadad, and J. Doorn, “Correcciones semánticas en métodos de estimación de completitud de modelos en lenguaje natural,” presented at the WER 2013 – 16th Workshop on Requirements Engineering, Montevideo, Uruguay, 2013, pp. 105 – 117.

Aplicación de patrones en el modelado de procesos de negocio

Marcelo Marciszack, Claudia Castro, Claudia Sánchez, Andrea Delgado

Dpto. Ingeniería en Sist. de Información/ Facultad Regional Córdoba/ Universidad Tecnológica Nacional

{ marciszack, ingclaudiacaastro, claudiasanchez001, andreafdelgado}@gmail.com

Resumen

El objetivo del presente trabajo, es la aplicación de patrones en el modelado de procesos de negocio que permita generar un modelo inicial; a partir del cual se deriven los requerimientos/casos de usos para su validación, creando Modelos Conceptuales de un sistema de información. El presente puede considerarse como una extensión al proyecto anterior "UTN1643 -Validación de Requerimientos a través de Modelos Conceptuales" (consolidado dentro del grupo GIDTSI dependiente del Departamento de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, Facultad Regional Córdoba). Desde lo metodológico, se apunta al uso de patrones de negocio, análisis y diseño, en el desarrollo de un Modelo Conceptual que permita un primer acercamiento inicial al nuevo sistema, permitiendo facilitar la validación y verificación de los requerimientos funcionales que contempla el sistema. Para que pueda ser factible, será necesario poder reusar los conocimientos y experiencias adquiridas en el desarrollo de sistemas anteriores, siendo ese conocimiento un conjunto de soluciones a las que denominaremos patrones de análisis y diseño.

También se propone ampliar el alcance y funcionalidad de una herramienta

denominada "SIAR" que permite gestionar y validar requerimientos de software, diseñada y construida en el mencionado proyecto anterior.

Palabras clave: *Patrones, Procesos, Negocio, Modelado de Negocio.*

Contexto

El presente trabajo se encuentra en el marco del proyecto "Implementación de Patrones en la Validación de Modelos Conceptuales", que se encuentra consolidado dentro de la línea de investigación de Sistemas de Información en el Dpto. de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, Facultad Regional Córdoba.

Dentro de esta línea de investigación, se busca plantear la interpretación del negocio, bajo estándares que faciliten o agilicen la comprensión de los mismos garantizando el entendimiento de los procesos de negocios para la posterior identificación de los requerimientos del software.

Introducción

Un modelo de negocio es una visión simplificada de un negocio, es una abstracción de cómo funciona. Ofrece una visión simplificada de la estructura de negocios que servirá de base para la

comunicación, mejoras o innovaciones; así como definir los sistemas de información sobre los requisitos que son necesarios para apoyar el negocio.

El modelado del negocio es la técnica por excelencia para alinear los desarrollos de software con las metas y objetivos de las empresas e instituciones. El modelado de negocios, y, más específicamente el modelado de procesos de negocio, es la forma idónea para comunicarnos con los usuarios de todos los niveles.

Contar con un modelo de negocio para modelar la arquitectura de software permite:

- Obtener un buen conocimiento de los requisitos del negocio sobre sus sistemas de apoyo.
- Contar con gran cantidad de información vital que aumenta la calidad del sistema de software.
- Utilizar el mismo lenguaje de modelado lo cual aumenta la trazabilidad entre los modelos. Esto significa que una función específica en el sistema de información se remonta a un requerimiento específico en el negocio y un cambio posterior en el modelo de negocio puede más fácilmente ser propagado al modelo de software.

Un objetivo adicional, es crear el sistema de software(s) que mejor soporte y ajuste el negocio, por lo tanto, el modelo de negocio es una base muy importante tanto para especificar los requisitos como para diseñar el software.

Patrones de procesos de negocios

¿Qué es un patrón?

En el ámbito del desarrollo de software, un patrón es una descripción de un problema y la solución, a la que se le da un nombre, y se

la puede aplicar a nuevos contextos, idealmente proporciona consejos sobre el modo de aplicarlo en distintas circunstancias [Larman, 2003].

Los Patrones de Procesos de Negocios (PPN) permiten modelar y rediseñar los procesos de negocio en una organización. Son **estructuras genéricas** que establecen en forma sistémica los **macroprocesos, procesos, subprocesos y actividades** que deben existir en cualquier organización para hacer posible su funcionamiento.

La estructura es jerárquica y entrega mayor detalle en los niveles más bajos de ella.

Cada tipo de macro proceso tiene, a su vez, **un patrón que define su estructura interna en término de los procesos que lo componen y sus relaciones**. El más conocido de éstos es el de la **cadena de valor de una organización**, que permite establecer que la estructura de ella es **igual** para manufactura, hospitales, administración de justicia, distribución, etc. Los patrones se detallan por descomposición jerárquica en varios niveles de detalle, definiendo subprocesos y actividades. En estos niveles de detalle aparecen **mejores prácticas** que recomiendan la manera más adecuada de ejecutar un proceso de negocio.

Los patrones del proceso de negocio están orientados a **modelar la estructura sistémica de un negocio**, enfatizando las relaciones entre los diferentes procesos que existen y cómo se puede optimizar la coordinación entre ellos, por medio de lógica de negocio bien diseñada y automatizada parcial o totalmente por medio de apoyo TI.

Modelando el negocio con patrones

Los patrones de modelos de negocio se clasifican en:

- Patrones de Reglas y Recursos
- Patrones Meta
- Patrones de Proceso

Los patrones de Recursos y Reglas:

proporcionan directrices para el modelado de las reglas y recursos en el ámbito empresarial. Todas las empresas tienen que lidiar con los productos y documentos, por lo tanto es uno de los patrones más importantes que aborda esta clasificación.

Patrones de Meta: los patrones de meta se usan precisamente en el modelado de meta. El modelado de objetivo es un tema muy crítico, un modelo de validar y verificar objetivo apoya todo el trabajo de modelado. Los modelos de objetivo afectan el modelado de todo el proceso: cómo el sistema se construye y cómo se utiliza.

Patrones de Procesos: Son los patrones de comportamiento y funcionales, cuya intención es aumentar la calidad en los modelos de flujo de trabajo y otros modelos orientados al proceso.

Utilizando la arquitectura del negocio para definir el software

La transferencia de la información del modelo de negocio para el modelo de software no es un proceso simple o automático. Desafortunadamente, no hay un mapeo uno-a-uno para este proceso, y no hay ningún algoritmo simple para traducir el modelo de negocio en un modelo de software.

Se trata de dos modelos diferentes, con diferentes propósitos. El modelo de negocio describe una empresa o una parte específica

de un negocio, y no todas las partes de la empresa se transmiten en los sistemas de software (por ejemplo, las personas, actividades manuales, y muchos objetivos del negocio). Tampoco todas las clases y objetos en el modelo de negocio se convierten directamente en las clases y objetos del software; lo cual hace a menudo un tanto confusa la labor.

Las conexiones entre las dos arquitecturas se pueden extraer en la definición de una arquitectura de software basada en la arquitectura del negocio, a través de un Diagrama de proceso de negocio y de procesos de desarrollo de software, considerando las conexiones y los recursos que utilizan o producen en los procesos.

Líneas de investigación y desarrollo

Para lograr los objetivos del presente trabajo de investigación, se perseguirá como método de investigación, el método científico, haciendo incapié en la validez del supuesto de trabajo tomando como base al grado de aceptación y utilización que se le añadirá a la herramienta para la implementación de Patrones en diferentes proyectos de transferencia que se realizan dentro del GIDTSI.

Para la realización del presente trabajo se describen de manera breve las siguientes líneas de investigación y desarrollo.

1) Investigación de patrones existentes para modelar el negocio. Se llevará a cabo una exhaustiva búsqueda bibliográfica de papers y documentación pertinente, como así también artículos relacionados al tema tratado en el presente proyecto.

- Se completará el estudio de las distintas estructuras y clases de patrones.
- Se analizará y evaluarán las características que definen a los patrones identificados mediante la facilidad de implementación y aplicación, el grado de representación y completitud en su definición.
- Se procederá a la aplicación de patrones en el Modelo de Negocio, validando su factibilidad y viabilidad en su uso.

II) Propuesta Metodológica

- Basándose en un proceso de negocio, se aplicarán patrones específicos asociados a las diferentes actividades de un negocio en particular. Se asociarán cada una de las actividades para lograr la construcción de un Modelo Conceptual basado en patrones.

III) Elección de Patrones

- Se caracterizará y definirán ámbitos de aplicación a cada uno de los tipos de Patrones identificados, de acuerdo a su funcionalidad.
- Para cada uno de los patrones con los que se trabaje, se definirán de manera clara y precisa cada una de las funcionalidades, el objetivo de cada patrón y los atributos que lo define. Se formalizará una heurística de asignación para la asociación de cada patrón a cada actividad de Negocio.

IV) Definición de plantillas de procesos de negocio

Se definirá la estructura de los templates para soportar a las plantillas para cada proceso de negocio asociado a cada uno de los patrones predefinidos.

Se estudiará el impacto que tiene la aplicación de los patrones en la determinación posterior de los requerimientos funcionales.

V) Trazabilidad de patrones

Formará parte de la definición del proceso, la individualización de los aspectos que dan origen a una funcionalidad específica, el proceso de negocio al que se encuentra vinculado y la forma en la que fue implementado el patrón.

VI) Adaptar y Ampliar la funcionalidad del SIAR

- Se realizará el rediseño de la herramienta SIAR para que pueda dar soporte la inclusión de diferentes patrones en el Modelado de Procesos de Negocio.
- Se diseñarán y desarrollarán diferentes prototipos basados en el rediseño realizado.
- Se realizará “testing” de los prototipos y se evaluarán los resultados obtenidos, a través de diferentes herramientas visuales que permitan la comprensión por parte de los usuarios del negocio bajo estudio.
- Se construirá una nueva versión herramienta que de soporte a la nueva metodología, y que permita gestionar patrones de negocio.
- Se documentará el proceso de construcción de la herramienta.

VII) Utilización y divulgación

- Puesta en operación y ejecución en el centro de desarrollo.
- Confeccionar artículos y seminarios, para validar con la comunidad científica el avance de la propuesta, y generar divulgaciones sobre el tema interna y externa.
- Definir, desarrollar y ejecutar plan de capacitación y asistencia a usuarios finales para transferencias.

Resultados y Objetivos

Se prevé la transferencia de resultados en primer lugar con el CIDS (Centro de

Investigación y Desarrollo de Sistemas), de la misma Facultad para su uso interno, y a posteriori con otras redes de investigación dentro del ámbito de la Ingeniería de Software.

Al mismo tiempo, y como contribución a la formación de los integrantes del proyecto, se elaboraran informes técnicos, como así también artículos para ser publicados en Congresos, Conferencias y reuniones Científicas tanto a nivel nacional como a nivel internacional.

Formación de Recursos Humanos

En el marco del actual proyecto de investigación, en el 2015, se doctoró en la Universidad de Vigo, Marcelo Marciszack bajo la dirección del Dr. Manuel Pérez Cota con tesis que abordó el tema de "Validación de especificaciones funcionales de esquemas conceptuales a través de abstracciones"; y obtuvo el título de Especialista en Ingeniería en Sistemas de Información en la U.T.N. F.R.C. de la Ing. Claudia Castro, cuyo tema de trabajo es "Un modelo conceptual para la obtención del sistema de Información a partir del Sistema del Negocio". Además se incorporan tres becarios alumnos de investigación, un becario Graduado BINID, que serán de ayuda en la recolección y manipulación de bibliografía, y colaborarán en el diseño y construcción de la herramienta informática. Al mismo tiempo, y como contribución a la formación de los integrantes del proyecto, se elaboraran informes técnicos, como así también artículos para ser publicados en Congresos, Conferencias y reuniones Científicas tanto a nivel nacional como a nivel internacional.

Referencias

[Carrasco 2010] Gestión de Procesos (La Participación es la clave). Juan Bravo Carrasco, Editorial Evolución S. A. 2010.

[Larman 2003] Larman, Craig. UML Y Patrones: Una introducción al análisis y diseño orientado a objetos y al proceso unificado. Segunda Edición. Pearson Educación, S.A., Madrid, 2003.

[Barros 1999] Dr. Oscar Barros. Patrones de Procesos de Gestión – Compartiendo Conocimiento para Aumentar la Productividad – Patrones de Proceso. La arquitectura general de un proceso. Documento de Trabajo Nro. 9, Departamento de Ingeniería. Universidad de Chile, enero 1999.

[Braude 2010] Braude, Eric J., and Michael Bernstein. Software Engineering: Modern Approaches, 2nd ed. Wiley, 2010.

[Foerster 2005] Foerster, Alexander; Engels, Gregor; Schattkowsky, Tim. Activity diagram patterns for modeling quality constraints in business processes. En Model Driven Engineering Languages and Systems. Springer Berlin Heidelberg, 2005. p. 2 - 16.

[Pressman 2010]: Ingeniería del software. Un enfoque práctico. 7ma. Edición. Pressman Roger S. Editorial Mc Graw Hill Educación. Impreso en México © 2010.

[Fowler97] Fowler, M., Analysis Patterns: Reusable Object Models, Addison - Wesley. - 1997.

[Sommerville 2011] Novena Edición en castellano del Libro Ingeniería de Software de Ian Sommerville Editado por Pearson Educación – México 2011 Versión impresa ISBN 978-607-32-0603-7

Aplicación de tecnologías semánticas a la Forensia Digital

ETAPA 1: Estudio y Diseño de una Ontología Semántica

Beatriz P. de Gallo^a, Horacio Leone^b

^aI.Es.I.Ing. /Facultad de Ingeniería, Universidad Católica de Salta
Campo Castañares S/N, Salta, Argentina

^bINGAR/ Facultad Regional Santa Fe UTN
Avellaneda 3657, Santa Fe, Argentina

^abgallo@ucasal.edu.ar, ^b{hleone}@santafe-conicet.gov.ar

Resumen

Considerando la vinculación interdisciplinaria de la Informática y las Ciencias Jurídicas, se propone trabajar en el área de la Forensia Digital, particularmente mediante el empleo de la ingeniería ontológica para estudiar un conjunto de problemas básicos que puedan ser tratados y expresados eventualmente a través de una ontología.

El trabajo de investigación propone generar un contexto de análisis de la evidencia digital que permita una visión integrada, sistemática y orientada al sujeto. Se propone circunscribir el contexto de aplicación y experimentación del tema en estudio, según criterios de alcance, profundidad, oportunidad y acceso a problemáticas reales de la Forensia Digital, identificando los problemas básicos al momento de realizar el análisis forense de la información digital. Esto permitió acotar el universo de estudio al análisis forense de correos electrónicos y en la propuesta de una ontología dirigida a generar un espacio de comunicación común entre todos los actores del proceso judicial (jueces, abogados, peritos, etc.), con la intención de que la información técnica proveniente del análisis forense de correos electrónicos se incorpore a la causa con

idénticas consideraciones sobre el valor probatoria que otras pruebas documentales.

Palabras clave:

Ontologías Semánticas, Forensia Digital

Contexto

En respuesta a la política institucional de definición de planes de desarrollo de la investigación en todas las unidades académicas con especificación de prioridades, la Facultad de Ingeniería de la UCASAL definió sus líneas de investigación prioritarias, atendiendo a la evolución de las actividades de investigación desarrolladas en la facultad, a la conformación de equipos docentes y la demanda del medio productivo local. Entre ellas, interesa la referida a “*Tecnologías Informáticas Aplicadas*” que sirven de marco de referencia para el GRUPO DE INVESTIGACIÓN EN INFORMÁTICA FORENSE, integrado por docentes de la carrera de Ingeniería en Informática de esa Casa de Altos Estudios.

Este proyecto se presentó al Consejo de Investigaciones de la UCASAL, y pasó las instancias de admisibilidad y evaluación por pares externos según lo

estipula la RR N° 1083/14 en el procedimiento de presentación de proyectos de investigación. Por RR N° 656/15 se aprueba el proyecto y su financiamiento por parte de la UCASAL.

En este contexto se desarrolla el proyecto que aquí se presenta.

Introducción

Desde hace más de una década que los diferentes estamentos de seguridad –tanto militares como judiciales y políticos- se preocupan por encarar la lucha contra el crimen desde la óptica tecnológica, es decir, con una mirada cada vez más preocupante sobre el uso de la tecnología para delinquir. En 2001 la Digital Forensic Research Conference (DFRWS) definió la “Forensia Digital” como “El uso de métodos científicamente derivados y provados a la preservación, recolección, validación, identificación, análisis, interpretación, documentación y presentación de la evidencia digital derivada de fuentes digitales para el propósito de facilitar o favorecer la reconstrucción de los hechos criminales o para la prevención de acciones no autorizadas que se estima como perjudiciales para operaciones planificadas”[1].

La Forensia Digital ha entrado en una crisis producto del impacto de dos elementos que marcan la época actual de la tecnología informática: la masividad de los datos y la multiplicidad de plataformas tecnológicas. Garfinkel [2] presenta varios desafíos, involucrando no solo los modelos de “visibilidad y búsqueda” que proponen las herramientas forenses de uso actual sino también la falta de integración de las estrategias (como la ingeniería reversa) con dichas herramientas para reducir tiempos y costos. Cita este autor como próximos desafíos a resolver:

- Diseño de las herramientas orientadas a la evidencia: usualmente las herramientas actuales se orientan a la búsqueda de elementos digitales (evidencia) pero no a la presentación, resumen o análisis de correlaciones entre los datos encontrados.

- Modelo de visibilidad, filtro e informe: las herramientas utilizan interfaces de comunicación con el experto forense que habitualmente no permiten establecer vínculos o relaciones de prioridad entre los datos encontrados. Incluso algunas herramientas se basan en algoritmos computacionales costosos en tiempo y pueden faltarle características de usabilidad para el usuario final. La automatización o generación de scripts para búsqueda y filtro no siempre resultan. Y se complica aún más ante el avance continuo de las tecnologías (procesamiento paralelo, virtualización, deep web, etc.)

- Problemas estructurales en las herramientas forenses: en muchos casos se recurre a software desarrollado para el contexto de negocios o para sistemas transaccionales y no responden exactamente a las necesidades puntuales de la búsqueda de evidencia digital. Ocurre lo mismo con tecnologías integradas, tales como la ingeniería reversa o las aplicaciones monolíticas.

- Abstracción y modularización: debido al volumen de datos que se procesan en la búsqueda de la evidencia digital, se requiere fijar estándares para la identificación, transmisión e intercambio de los datos; igualmente es importante generar arquitecturas de procesamiento que superen los conflictos del software abierto y propietario.

- Enfoque en la identidad del individuo: tomando como atributos todos aquellos datos que puedan generar una “imagen” de la persona (datos de

identificación, datos bancarios, correos, vínculos de las redes sociales, etc.).

En el contexto forense, es de suma importancia vincular los datos a partir del significado de cada cosa. No se trata solo de “encontrar la evidencia digital”, sino de interpretarla en el contexto de la situación, vinculándola con el resto de los componentes de la investigación (pruebas físicas, interrogatorios, marco legal y procedimental del caso, etc.). De modo que es indispensable avanzar en la forensia digital desde la óptica de la semántica –como elemento vinculante de todos los componentes del sistema- así como desde un marco referencial que pueda interpretarlo –una ontología-.

Si bien la definición más referenciada en la literatura es la de Gruber [3] “una ontología es una especificación explícita de una conceptualización”, vale detallar un poco más el concepto, tomando lo dicho por Reuver et al.[4] “Una ontología es la descripción conceptual y terminológica de un conocimiento compartido acerca de un dominio específico. Dejando de lado la formalización e interoperabilidad de aplicaciones, esto no es más que la principal competencia del término: hacer mejoras en la comunicación utilizando un mismo sistema en lo terminológico y conceptual”.

Líneas de Investigación, Desarrollo e Innovación

El Plan de trabajo consiste en profundizar el análisis del dominio de la forensia digital, sus requerimientos de información y estudiar los casos vinculados al análisis forense de correos electrónicos. Se identificarán delitos de interés social tomando como organismo público de investigación de delitos de referencia al Ministerio Público de la

Provincia de Salta y se tratará la naturaleza jurídica de la prueba digital –y en particular- del correo electrónico.

A partir de los requerimientos identificados, se formalizarán las debilidades de las arquitecturas y modelos desarrollados hasta el presente y las razones de las mismas, y se comenzará en el diseño de la ontología que represente el conocimiento requerido en el caso seleccionado.

Se seleccionó METHONTOLOGY como metodología de trabajo para la definición de la ontología, y se prevé el desarrollo de las etapas básicas (especificación de requerimientos, conceptualización, implementación y evaluación). A fin de evaluar la ontología en una aplicación se procederá a la implementación de una herramienta a nivel de prototipo.

En resumen, para el desarrollo de este plan se proponen las siguientes actividades: Búsqueda bibliográfica, estudio del estado del arte; definición del contexto de aplicación y experimentación; estudio de factibilidad técnica y operativa de aplicación; adquisición del conocimiento; especificación de usos y usuarios de la ontología; conceptualización y estructuración del conocimiento en modelos significativos; formalización del modelo; implementación mediante modelos computables y validación y ajuste de la ontología.

Resultados y Objetivos

La Figura 1 resume el grado de avance en la conceptualización de la ontología. Se describen los principales conceptos, atributos de instancias y las relaciones más notorias.

considerando que durante el proceso de transmisión el correo consta de sucesivas ocurrencias que se almacenan en servidores transitorios hasta llegar a destino.

Formación de Recursos Humanos

La RF N° 153/12 de la Facultad de Ingeniería de la UCASAL avala la conformación de grupos de investigación en el ámbito de esa unidad académica, destinados a fomentar la asociación de los investigadores en conjuntos que sustenten un programa estable para favorecer el trabajo de investigación multidisciplinario, optimizando así el uso de los recursos, facilitando la inserción en el medio socio-productivo, la constitución de redes y aumentando la competitividad de los investigadores en la captación de recursos desde agencias nacionales e internacionales. De esta acción surge el Grupo de Informática Forense, al cual pertenecen los investigadores del presente proyecto.

El equipo de trabajo del proyecto está conformado por el Dr. Horacio Leone, director, MBA Ing. H. Beatriz P. de Gallo, co-directora del proyecto, el Ing. Esteban Rivetti y la Esp. Abog. Maria Isabel Rodriguez Virgili.

El proyecto prevé además la incorporación de alumnos investigadores: los alumnos Esteban Rivetti (luego incorporado al equipo de investigación), Natalia Farfán y Ebaneo Kao desarrollaron sus trabajos de finalización de carrera en temáticas vinculadas a esta línea de investigación.

Es intención que este proyecto de investigación actúe como marco de contención para el desarrollo de un trabajo de tesis doctoral, a cargo de la Ing. H. Beatriz P. de Gallo, actualmente doctorando en la carrera de Doctorado en Ingeniería Mención en Ingeniería en

Sistemas de Información, de la Universidad Tecnológica Nacional – Facultad Regional Santa Fe.

Referencias

[1] DFRWS TECHNICAL REPORT.2001. A Road Map for Digital Forensic Research, <http://www.dfrws.org/2001/dfrws-rm-final.pdf> Página vigente al 15/11/2013

[2] Garfinkel, Simson L.2010. Digital forensics research: The next 10 years, <http://dfrws.org/2010/proceedings/2010-308.pdf> página vigente al 15/11/2013

[3] Gruber, Thomas R. 1993. A Translation Approach to Portable Ontology Specifications. Knowledge Systems Laboratory. Technical Report KSL 92-71

[4] Reuver, Mark de. y Haaker Timber, 2009, Designing viable business models for context-aware mobile services. Elsevier, Volume 26, Issue 3, Telematics and Informatics, Pages 240–248 (August 2009)

[5] Beatriz P. de Gallo, Marcela Vegetti, Horacio Leone, “Ontología para el Análisis Forense de Correo Electrónico”, CoNaIISI 2014 - ISSN: 2346-9927 - Página 1008-1018

[6] Gallo Beatriz P. de, Vegetti Marcela, Leone Horacio, “Población de ontologías con datos no estructurados utilizando herramientas de minería de datos”, CoNaIISI 2015 Actas del 3° Congreso Nacional de Ingeniería Informática/Sistemas de Información, Buenos Aires, Argentina, ISBN: 978-987-1896-47-9, 2015

[7] Rivetti E., Parra H.B., “Verificación de la trazabilidad de un correo electrónico mediante un caso ejemplo”, Cuadernos de Ingeniería 2015., Número 9 del 2015. ISSN 2422-6572 (On line), ISSN 2422-6564, in press.

[8] Gallo Beatriz P. de, Vegetti Marcela, Leone Horacio, “Avances en la Construcción de una Ontología para el Análisis Forense de Correo Electrónico”, presentado al VI CIHDI (Congreso Iberoamericano de Investigadores y Docentes de Derecho e Informática), en revisión.

Gestión del Conocimiento Abierto

María Valeria Poliche¹, Carola Victoria Flores¹, Marcela Soledad Molina¹,
María Belén Leguizamón¹, María Alejandra Barrera¹

1) Departamento Informática. Facultad de Tecnología y Ciencias Aplicadas.
Universidad Nacional de Catamarca, Argentina.

Maximio Victoria 35 (4700), Catamarca, Argentina. Tel: +54 (383) 4435112.

e-mail: {vpoliche, carolaflores, marcela_molina}@tecno.unca.edu.ar; marita.latingeo@gmail.com

Resumen

En los últimos años ha comenzado una verdadera revolución en materia de acceso a la información pública por parte de los ciudadanos. A esta revolución se le ha denominado “Datos Abiertos”, la cual consiste en poner a disposición de la sociedad los datos de interés común de la ciudadanía para que de cualquier manera, estos puedan desarrollar una nueva idea o aplicación que entregue nuevos datos, conocimientos u servicios que el gobierno no es capaz de entregar. El uso del conocimiento para una mejora de las estructuras organizativas y sociales ha dado lugar a un gran abanico de herramientas tecnológicas cuya finalidad es soportar estas estructuras y facilitar los flujos de conocimiento entre los agentes que las componen.

Esta línea de investigación aborda el tema de “Gestión del Conocimiento Abierto” en las organizaciones para gestionar en forma eficaz el conocimiento. Es también de gran importancia en las instituciones educativas tener y potenciar buenos sistemas de gestión del conocimiento abiertos a la innovación y la actualización continua, que cuenten con la participación activa de todos sus miembros, para lo

cual es clave el uso común de los nuevos medios sociales digitales.

Palabras clave: gestión del conocimiento, conocimiento abierto, TICs.

Contexto

Esta investigación se enmarca en el proyecto "TIC al servicio del dato abierto: situación actual, conceptualización e iniciativas de apertura de información pública”, y que tiene como objetivo general analizar la situación actual de las TICs al servicio del Dato Abierto contextualizando el análisis de los datos en iniciativas de apertura de información pública en Argentina y la región; produciendo una contribución efectiva de aportes concretos que impliquen nuevas propuestas TIC que asistan al desarrollo de software específico, permitiendo y alentando la reutilización de datos que deriven en la creación de aplicaciones y servicios para información primaria y accesible.

Dicho proyecto fue presentado en la Convocatoria de Proyectos I+D 2016 y se encuentra en la etapa de aprobación por parte de la Secretaria de Ciencia y

Tecnología de la Universidad Nacional de Catamarca(SECyT).

Cabe aclarar que los temas que se tratan en la línea de investigación se vienen trabajando desde el año 2013 con el proyecto “La gestión del conocimiento en organizaciones públicas y privadas de nuestro medio” Código: 02/1705 y que tuvo una duración de 3 (tres) años.

Introducción

Internet, las tecnologías y la cultura digital han sido claves en la actual combinación de elementos que refundan la idea de gobierno abierto. Por un lado, porque el entorno digital en nuestros días constituye la infraestructura de base que facilita y hace posible un acceso más fluido a la información y la Gestión del Conocimiento(GC); y por el otro, los valores subyacentes en los patrones de interacción que promueve la cultura digital: prácticas abiertas, transparentes, participativas y colaborativas. Todo ello es condición necesaria pero no suficiente para la construcción de gobiernos abiertos.

Las administraciones públicas disponen de muchos conjuntos de datos que, en buena parte pueden ser abiertos a la sociedad sin que implique ningún problema de privacidad. Estos datos pueden ser reutilizados para otras finalidades, lo cual constituye el principal objetivo de los procesos de apertura de la información pública.

La República Argentina es miembro de la Alianza para el Gobierno Abierto desde el año 2012, la cual es promotora de los principios de transparencia, colaboración y participación a nivel internacional y cuenta con la apertura de datos [1] que se rige en el siguiente marco legal:

- Plan Nacional de Apertura de Datos: Decreto 117/2016, reglamenta plazos

para la publicación de los datos en el Portal de Datos[1] como también para detallar los activos de datos de los distintos organismos de la Nación y su cronograma de publicación[2].

- Sistema Nacional de Datos Públicos: Resolución 538/2013, la cual reglamenta la creación del programa SINDAP, Sistema Nacional de Datos Públicos[3].
- Régimen de Acceso a la Información Pública: Decreto 1172/2003 que establece el Reglamento General del Acceso a la Información Pública para el Poder Ejecutivo Nacional[4].

El caso de Buenos Aires, la “Ley Acceso a la Información” N° 104/98 sancionada por la Legislatura de la Ciudad de Buenos Aires, consagra el derecho de los ciudadanos a solicitar y recibir información adecuada de los órganos de la administración pública porteña[5], quien posee un portal “Buenos Aires Data” donde se tiene acceso a los datos[6]. En cuanto a portales de datos abiertos, el software más acreditado que los soporta se denomina CKAN [7], fue desarrollado por la organización sin fines de lucro Open Knowledge Foundation[8], y en la actualidad más de 40 iniciativas de datos abiertos lo utilizan.

Por su parte, en las universidades poco a poco van materializándose iniciativas de apertura en las que se explota el potencial que el movimiento open puede aportar a estas instituciones de educación superior[9]. Para divulgar el conocimiento científico generados por la investigación se aprobó la Ley 26899 de “Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos” [10].

Como antecedente en este contexto se encuentra a la Universidad Nacional de Córdoba (UNC), Argentina quien plantea un modelo para institucionalizar el acceso

abierto en las universidades mediante una “Oficina de conocimiento abierto”, definida para asesorar y capacitar a los miembros que integran los diferentes proyectos sobre acceso abierto iniciados dentro del ámbito universitario[11].

Las organizaciones no solo deben disponer de medios tecnológicos para la generación, síntesis y transmisión del conocimiento, sino que deben existir otros sistemas que faciliten el flujo del conocimiento. Así, aparece la necesidad ineludible de proporcionar a las organizaciones los conocimientos que necesitan, donde los necesiten, cómo los necesiten y cuándo los necesiten. La GC, como su propio nombre indica, atiende a la dirección del recurso conocimiento y se define como aquella función que planifica, coordina y controla los flujos de éste que se producen en la empresa en relación con sus actividades y con su entorno para crear unas competencias esenciales [12][13]. Este proceso exige a su vez la apuesta empresarial por una serie de instrumentos de gestión que promueven la adecuada creación, acumulación y utilización del conocimiento, recurso empresarial por excelencia.

Como consecuencia las organizaciones que deciden implementar tecnologías relacionadas con la GC deben realizar cambios organizativos y, en muchos casos, cambios de cultura para conseguir que el uso de estas herramientas tecnológicas acompañados de otros sistemas no tecnológicos lleve a una mejora de los procesos de la organización [14][15][16].

El objetivo de la *Gestión del Conocimiento Abierto* es revisar el concepto y la extensión del movimiento o de la filosofía del conocimiento abierto en las universidades y e instituciones del medio. Este modelo consta de cuatro áreas: software libre, contenidos

educativos abiertos y difusión de la cultura, la ciencia y la innovación. Para cada una de las cuatro áreas mencionadas anteriormente se tratará de mostrar la forma en que están promoviendo la publicación y transmisión de la información en un entorno abierto, sin restricciones y favorecer la difusión de conocimientos en todos los campos[17][18].

El Conocimiento Abierto es un planteamiento que, aunque haya una cierta controversia crece sin cesar y las organizaciones internacionales y los gobiernos están adoptando gradualmente el conocimiento abierto como una forma de compartir los avances científicos con la sociedad y como una medida de la cooperación internacional para favorecer el desarrollo en los países del tercer mundo.

Líneas de Investigación, Desarrollo e Innovación

La línea general de investigación que se aborda es “*la Gestión del Conocimiento Abierto*”. Se pretende promover el estudio, formación, investigación y divulgación de la Gestión del Conocimiento Abierto en la comunidad académica y empresarial, identificando posibilidades de mejora e innovación proponiendo técnicas, metodologías y el desarrollo de herramientas que generen nuevos conocimientos a partir de fuentes de datos combinadas y patrones en grandes volúmenes de datos.

Por otra parte se pretende conocer las experiencias de implantaciones exitosas de herramientas TIC para garantizar la correcta aplicación del Dato Abierto.

Objetivos

Para el desarrollo de la línea de investigación el equipo de trabajo se ha planteado los siguientes objetivos:

- Estudiar tecnologías de información que den soporte a la GC de datos abiertos.
- Comprender las distintas estrategias de GC para identificar, recolectar y organizar el conocimiento existente en las organizaciones y así facilitar la creación de nuevo conocimiento.
- Lograr un mayor entendimiento de la GC y de las TIC que den soporte a esta.
- Contribuir a la formación y capacitación de docentes investigadores en las distintas estrategias de cambio y herramientas TIC.
- Concientizar a la comunidad universitaria de que la investigación financiada con fondos públicos debe gozar de amplia difusión a través de la publicación en acceso abierto de datos y documentos científicos.
- Promover la creación y difusión de conocimientos producidos colaborativamente entre distintos actores socioeducativos.

Formación de Recursos Humanos

El equipo de trabajo de la línea de investigación se encuentra conformado por: una directora, una co-directora, tres docentes investigadores.

Cuatro de los docentes que integran el proyecto están categorizados en el Programa de Incentivos de la Secretaría de Políticas Universitarias (SPU) perteneciente al Ministerio de Educación de la Nación Argentina.

Dos miembros del equipo de trabajo han finalizado el cursado y se encuentran desarrollando la tesis de la Maestría de Informática de la Facultad de Ingeniería de la Universidad del Norte Santo Tomás de Aquino. San Miguel de Tucumán.

En el marco de esta investigación se desarrolló y terminó en el año 2015 una tesis de maestría en Ingeniería de Software de la Universidad Nacional de San Luis (UNSL) denominada “Gestión del Capital Intelectual en el ámbito universitario: sobre la función de investigación”, la cual presenta un modelo para la medición del capital intelectual para la investigación universitaria, tomando como caso de estudio la SECyT.

En lo que se refiere al desarrollo de tesis de grado:

- Se asesoró una tesis de grado para la carrera de Licenciatura en Sistemas de Información, de la Universidad Nacional de Catamarca (UNCa).
- Se encuentran en proceso de desarrollo dos trabajos de tesis para la carrera Ingeniería en Informática de la UNCa.

Se pretende acrecentar la formación continua de los profesionales investigadores y mantener líneas de trabajo que sirvan a nuestro medio para proveer nuevas estrategias de administración de su conocimiento.

Referencias

- [1] Datos Argentina. Ministerio de Modernización de la República Argentina. Portal de datos abiertos del Gobierno de la República Argentina. <http://www.datos.gob.ar/>
- [2] Decreto 117/2016: Plan Nacional de Apertura de Datos. Ministerio de Modernización. Buenos Aires.,

- 12/01/2016.
- [3] Resolución 538/2013: Programa Sistema Nacional de Datos Públicos. Jefatura de Gabinete de Ministros. Buenos Aires, 18/7/2013.
- [4] Decreto 1172/2003: Régimen de Acceso a la Información Pública. Buenos Aires, 3/12/2003.
- [5] Buenos Aires Data. <http://data.buenosaires.gob.ar/>
- [6] Ley N° 104 Acceso a la Información Pública: regula el Acceso a la Información Pública en el ámbito de la Ciudad de Buenos Aires. 1998.
- [7] CKAN: The open source data portal software. <http://ckan.org/>.
- [8] Open Knowledge Foundation, <https://okfn.org/>.
- [9] García Peñalvo, F. J.; C. García de Figuerola, J. A. Merlo Vega. Open Knowledge Management in Higher Education. 2010 [en línea] Disponible en: <http://hdl.handle.net/10366/121874>.
- [10] Ley 26899: Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos. 13 de noviembre de 2013. [en línea]. Disponible en: <http://repositorios.mincyt.gob.ar/recursos.php>.
- [11] Oficina de conocimiento abierto: un modelo para institucionalizar el acceso abierto en las universidades. <https://rdu.unc.edu.ar/handle/11086/1367>.
- [12] Pérez-Montoro Gutiérrez, M. Gestión del conocimiento en las organizaciones: fundamentos, metodología y praxis. Gijón: Trea, 2008. ISBN 978-84-9704-376-2.
- [13] Thomas Otter y M. Cortez. Gestión del Conocimiento: Conceptos, Ideas, Herramientas. Diciembre 2013. Disponible en: http://www.kas.de/wf/doc/kas_21114-1522-4-30.pdf?101112184944.
- [14] GRAU, América (2002). Herramientas de Gestión del Conocimiento [en línea]. Fundación Iberoamericana del conocimiento. Disponible en: <http://www.gestiondelconocimiento.com/articulos.php>.
- [15] Carballo, Roberto. Innovación y gestión del conocimiento. Ediciones Díaz de Santos, 2006.
- [16] Ortiz, Angel Luis Arbonies. Conocimiento para innovar. Cómo evitar la miopía en la gestión del conocimiento. Ediciones Díaz de Santos, 2006.
- [17] González-Sánchez, Rocío, & Fernando E. García-Muiña. Innovación abierta: Un modelo preliminar desde la gestión del conocimiento. *Intangible capital* 7.1 (2011): 82-115. ISSN1697-9818
- [18] DAHLANDER, L.; MAGNUSSON, M. G. (2005). Relationships between open source software companies and communities. *Research Policy*, 34(4): 481-491.

Aseguramiento de la calidad en productos, procesos de software y procesos de gestión para la mejora de las sociedades del conocimiento

Esponda Silvia ⁽¹⁾, Pasini Ariel ⁽¹⁾, Boracchia Marcos ⁽¹⁾,
Díaz Delfina ⁽¹⁾, Calabrese Julieta ⁽¹⁾, Pesado Patricia ⁽¹⁾, Estevez Elsa ⁽²⁾

⁽¹⁾Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 120 La Plata Buenos Aires

⁽²⁾Universidad de las Naciones Unidas (UNU-EGOV) – Guimarães - Portugal

(sesponda, apasini, marcosb, ddiaz, jcalabrese, ppesado) @lidi.info.unlp.edu.ar, estevez@unu.edu

Resumen

La gestión de la calidad se impone en las organizaciones por la importancia que alcanza en diferentes aristas. A nivel de sus productos, pues permite establecer la calidad lograda y las características presentes en los mismos y a nivel organización, pues se ocupa de establecer un marco de procesos y estándares que permiten definir una mejora.

En particular, interesa aplicar la gestión de la calidad para obtener una mejora en los procesos de gestión universitaria.

El III-LIDI posee un grupo dedicado a la investigación y desarrollo de propuestas en temas relacionados con la mejora de los procesos de gestión y el aseguramiento de la calidad en procesos de gestión y productos de software.

Palabras Claves

Ingeniería de Software – Calidad del proceso – Calidad del producto - Normas de Calidad – ISO

Contexto

El Instituto de Investigación en Informática LIDI mantiene desde el año 2004 una línea de investigación y desarrollo en calidad, a través de proyectos acreditados por el Ministerio de Educación de la Nación,

apoyados por la Facultad de Informática y otros organismos nacionales e internacionales.

Esta línea de investigación se enmarca en el subproyecto– “Mejora de Procesos en el desarrollo de Sistemas de Software y en Procesos de Gestión. Experiencias en PyMEs” del proyecto “Tecnología para Sistemas de Software Distribuidos. Calidad en Sistemas y Procesos. Escenarios educativos mediados por TICs” (2014-2017), acreditado por el Ministerio de Educación de la Nación.

La Facultad de Informática subsidia un proyecto para el estudio de normas utilizadas en la certificación de productos que integran hardware y software denominado “Evaluación de ambientes de Educación a Distancia basado en la ISO/IEC 25000 (2do año)” (2016)

El Instituto posee diversos acuerdos de cooperación con varias Universidades de Argentina y del exterior, en particular con la Universidad de Naciones Unidas. Por otro lado, mantiene acuerdos con empresas privadas del sector, interesadas en mejorar sus procesos de desarrollo aplicando mejoras.

Se participa en iniciativas como el Programa IberoTIC de intercambio de

Profesores y Alumnos de Doctorado en el área de Informática.

Introducción

El uso de diferentes tecnologías de información y comunicación (TICs) ha modificado el desarrollo de un gran número de actividades cotidianas, principalmente en las relaciones sociales, culturales y económicas. El gran volumen de información registrada constantemente por el uso de las herramientas de las TICs, dan origen a la "sociedad de la información" donde todos pueden crear, acceder, utilizar y compartir información y el conocimiento, en donde se desarrolle y se mejore la calidad de vida". Un paso más evolucionado de la información, es el análisis de la misma para obtener conocimiento, creando el concepto de sociedades del conocimiento como una conceptualización de las innovaciones de las TICs para transferir la información para el desarrollo de actividades de las sociedades modernas.

La calidad de los procesos, que acompañan el crecimiento de las TIC es fundamental para obtener información de calidad, que es la entrada a las denominadas sociedades de conocimiento.

La incorporación de mejores prácticas en organizaciones desarrolladoras de software, tanto a nivel de realización de un producto, definición del proceso, o gestión, dejó de ser una aspiración para ser una necesidad para aumentar la competitividad de las mismas, ya sea para la obtención de una certificación y poder acceder a beneficios impositivos que promueve la denominada *Ley de Software* [9], o bien, con el objetivo de lograr una mejora a través del uso de buenas prácticas. La aplicación de estas prácticas en un proyecto, llevan hacia la definición de procesos concretos que facilitan la previsibilidad de la organización al momento de encarar nuevos proyectos [1].

El uso de normas y estándares para el logro del objetivo, es cada vez más común. Habitualmente los términos estándar y norma se utilizan como sinónimos, pero hay una diferencia importante que debemos tener en cuenta, el término **norma**, se define como, "*regla que se debe seguir o a que se deben ajustar las conductas, tareas, actividades, etc.*" mientras que el término **estándar** determina "*que sirve como tipo, modelo, norma, patrón o referencia*". El último término, hace referencia a elementos de carácter opcional mientras que la norma lo determina como obligatorio.

Las certificaciones se llevan a cabo bajo los requisitos que solicitan las normas, mientras que el cumplimiento de los requisitos puede ser asistido por modelos o estándares que facilitan su cumplimiento.

ISO y IEC son los principales productores de normas relacionadas con el software con reconocimiento internacional, IRAM es el único ente nacional con la capacidad de darle reconocimiento nacional a las normas publicadas por ISO o IEC.

Dentro de la línea de la investigación del proyecto, se destacan los siguientes ejes principales:

1- *Calidad en productos.*

La calidad de los productos de software se ha convertido en uno de los principales objetivos estratégicos de las organizaciones con la idea de lograr la competitividad que se espera en el mercado.

A partir de esta acepción, el Instituto ha iniciado una línea de investigación tendiente a allanar el camino de dichas organizaciones en el logro de esos objetivos.

En este sentido, el Instituto ha estudiado la norma IRAM – ISO/IEC 25000, que posibilita la certificación de los productos de software. Dicha norma define un conjunto de criterios (funcionalidad,

rendimiento, usabilidad, fiabilidad, seguridad, entre otros) para la evaluación de cualquier producto de software. Dichos criterios pueden ser evaluados por separado según el objetivo de la evaluación a realizar. [2,3]

El Instituto trabaja en la definición de una herramienta que permita, en base a los objetivos, establecer las características que serán evaluadas para un producto en particular, con qué métricas e indicadores. Esta herramienta está pensada para uso también en entornos virtuales de enseñanza y aprendizaje (EVEA), ampliando la gama de productos.

Los EVEA son aplicaciones informáticas que funcionan vía Web, diseñados para facilitar la comunicación pedagógica entre los participantes en un proceso educativo.

2- Mejora de los servicios de gobiernos digitales en unidades académicas de nivel universitario

En el marco de las mejoras de la calidad de los servicios institucionales, se inició una evaluación de la calidad de los servicios que brindan las unidades académicas (UA) con autonomía para definir sus procesos académicos y de gestión. Se definió un modelo, que toma como base el nivel de automatización de los procesos de gobierno, definido en [6], adaptados al ámbito universitario: **Emergente:** Una presencia formal en la web es establecida a través de sitios web aislados, publicando información estática con algunos medios de contacto oficiales. Esta fase permite a los receptores acceder a información sin necesidad de asistir a la unidad académica. **Mejorado:** Los sitios web son regularmente actualizados y contienen información dinámica, como publicaciones, legislaciones y boletines informativos. Se proveen enlaces entre sitios webs, especialmente entre sitios de la UA. Se provee información en las redes sociales. Se proveen servicios de búsquedas.

Interactivo: Provee acceso en línea a una gran variedad de servicios disponibles a los

receptores a través de sus sitios web. También se proveen servicios más avanzados, como por ejemplo, correos electrónicos a funcionarios, descarga y presentación de formularios de solicitud, y búsquedas en bases de datos especializadas. **Transaccional:** Se pueden ejecutar transacciones completas y seguras a través de un sitio web, como, por ejemplo, inscripción a materias o una solicitud de aulas. Un receptor puede completar toda la transacción electrónicamente. (Sin asistir a la UA en todo el proceso). **Integrado:** Un único punto de acceso al servicio. El servicio es complejo, intervienen varios actores de diferentes dependencias (internas o externa de la UA) y pasa por diferentes estados, se realiza todo de forma on-line sin asistir a la UA en ninguna instancia.

El primer paso de la evaluación consiste en definir la estructura de la UA, su comunidad, y seleccionar de los servicios que brinda la UA más representativos para cada clase de ciudadano (alumno, docente, etc.). Luego se realizará la evaluación sobre la prestación del servicio desde el punto de vista de la UA y una evaluación de satisfacción de los miembros de la comunidad con el uso de los servicios.

3- Mejora en los procesos de gestión de la Facultad de Informática

El Área de Certificaciones de Calidad de la Facultad de Informática coordina con este proyecto, con el objeto de analizar, definir y establecer un plan a ser aplicado a distintos procesos de la Gestión Universitaria.

Desde el año 2011, la Facultad de Informática ha iniciado el camino hacia la certificación de un proceso de ejecución anual, el curso de Nivelación a Distancia para el Pre Ingreso, logrando en el año 2012 la certificación IRAM-ISO9001:2008 del Sistema de Gestión de Calidad (SGC) del "Diseño y realización del curso de

Nivelación a Distancia para el Pre Ingreso a la Facultad de Informática". Actualmente el grupo ha obtenido la Re-certificación, garantizando así el nivel de la prestación del servicio por tres años. Esta tarea es desarrollada por los miembros del III-LIDI. Recientemente se realizó la primer auditoría interna del SGC del área de Concursos Docentes, obteniendo los resultados esperados, y dentro de las próximas semana el IRAM realizará las auditorías de certificación .[7]

4- Asesoramientos y consultorías

Dada la experiencia desarrollada por el grupo, existen numerosos convenios de consultoría con organismos estatales provinciales e instituciones privadas.

Líneas de investigación y desarrollo

- Análisis y estudio de normas y modelos orientados a la Calidad del Proceso, como por ejemplo IRAM - ISO/IEC 15504, IRAM - ISO/IEC 12207, ISO 9001 e ISO/IEC 90003. [4, 5, 7, 8,10]
- Análisis y estudio de normas y modelos orientados a la Calidad del Producto, como por ejemplo IRAM - ISO/IEC 25000
- Evaluación de procesos en organismos públicos y privados según los requisitos de ISO 9001.
- Análisis, discusión y estudio de normas de calidad relacionadas con certificación de servicios gubernamentales.
- Evaluación de madurez de los servicios de gobierno digital de una unidad académica.

Resultados obtenidos / esperados

Durante el año 2015 se realizaron las siguientes actividades:

- Capacitación y desarrollo de los documentos básicos de gestión de la calidad de productos.

- Aplicación de métricas para la evaluación de calidad de producto.

- Se avanza en una herramienta orientada a la mejora de procesos y gamificación.

- Se avanza en la evaluación de diferentes de unidades académicas para obtener una visión general del estado de madurez de los servicios de gobierno digital universitario.

- Mantenimiento de los procesos certificados ISO 9001:2008 del Pre-Ingreso a Distancia de la Facultad de Informática. Alcanzando los tres años de certificación y obteniendo la Re-Certificación por tres años más.

- Aplicación del requisito de un SGC al área concursos docentes para su próxima certificación.

- Análisis y preparación de posibles nuevos procesos de la Facultad, a ser certificados.

- Acciones de consultorías y asesoramiento en organismos públicos y privados

- Realización de curso de actualización de la Norma ISO 9001:2015 en el Instituto Argentino de Normalización y Certificación (IRAM).

- Formación de recursos humanos de grado y postgrado.

Formación de recursos humanos

- Capacitación de los miembros del proyecto a través de diversos cursos del Instituto Argentino de Normalización y Certificación (IRAM).

- Se desarrollan tesis de doctorado y tesinas de grado en el área.

- Los integrantes de esta línea de Investigación participan en el dictado de asignaturas/cursos de grado/postgrado en la Facultad de Informática de la UNLP y en otras universidades del país. En particular, en la UNLP, se dicta la asignatura "Calidad de Sistemas de Software en Pequeñas y Medianas Empresas".

Referencias

- [1] Pantaleo Guillermo, Calidad en el desarrollo de software, Editorial Alfaomega, ISBN 978-987-1609-23-9 Año 2011
- [2] ISO/IEC 9126, Software engineering -- Product quality
- [3] ISO/IEC 25000, Software Engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Guide to SQuaRE
- [4] ISO/IEC .15504: 2004 Information Technology – Process assessment
- [5] ISO/IEC 12207, Systems and software engineering -- Software life cycle processes
- [6] Estevez, E. (UNS) 2009. Servicios de Mensajería Programable para Gobierno Electrónico. (2009).
- [7] IRAM-ISO. 9001:2008, “Sistema de gestión de la calidad. Requisitos” Segunda Edición 2008 IRAM
- [8] ISO/IEC 90003:2004, “Software engineering — Guidelines for the application of ISO 9001:2000 to computer” 2004, ISO
- [9] Ley 25.922 - LEY DE PROMOCION DE LA INDUSTRIA DEL SOFTWARE - http://www.mincyt.gob.ar/multimedia/archivo/archivos/ley_25922_3.pdf
- [10] ISO/IEC 29110:2011, “Software engineering - Lifecycle profiles for Very Small Entities (VSEs)” 2011, ISO

MoVer-T: Entorno de videojuego en Realidad Virtual Telecontrolado para Rehabilitación Motriz

Cintia Ferrarini¹, Mónica González², Sergio Zapata¹, Luís Olgún¹, Emilio Ormeño¹

¹Instituto de Informática / Facultad de Ciencias Exactas, Físicas y Naturales / Universidad Nacional de San Juan.

²Departamento de Informática/Facultad de Ciencias Exactas, Físicas y Naturales / Universidad Nacional de San Juan.

¹ {eormeno, ferrarini, lolguin, szapata}@iinfo.unsj.edu.ar,

² gonzalez.monicagilda@gmail.com

Resumen

Los videojuegos aplicados al ámbito de la discapacidad cuentan con un gran potencial como recursos educativos, de ocio, y de tratamiento de rehabilitación motriz. En este último caso, expertos consideran que la ejercitación para rehabilitación mediante videojuegos en realidad virtual (VR-Games) realizada en la propia casa de los pacientes, constituye el futuro de estas terapias. Así, la realidad virtual en entornos telecontrolados se convierte en una herramienta valiosa debido a su capacidad de representar situaciones de la vida real en un contexto de total seguridad; además de facilitar el acceso a un tratamiento continuo y controlado a aquellos pacientes que poseen complicaciones de traslado, tanto de distancia como de tiempo. Este proyecto tiene como objetivo desarrollar un videojuego de realidad virtual multijugador en línea (Multiplayer Online Virtual Reality Game, o MOVR-Game) con dos aplicaciones interconectadas: una para los pacientes que se ejecutará en los lentes de realidad virtual; y otra para el

profesional de la salud, quien desde su consultorio podrá realizar la gestión, seguimiento y adaptación de las sesiones de juego de cada uno de sus pacientes. La tecnología desarrollada tendrá como destinatarios a los pacientes de la Fundación Abril, organización de la salud que brinda rehabilitación a personas con discapacidad entre otras actividades.

Palabras clave: videojuegos en realidad virtual, rehabilitación motriz telecontrolada, tecnología para la salud.

Contexto

El Proyecto MoVer-T: Entorno de videojuego en Realidad Virtual Telecontrolado para Rehabilitación Motriz, por sus características se enmarca en Proyectos de “Desarrollo Tecnológico y Social (PDTS), correspondiente a la convocatoria 2016-2017 y en la línea de investigación “Videojuegos para a la salud”. En esta línea el equipo de investigadores viene trabajando desde el año 2013, en el proyecto de investigación denominado

“Impulso a la Producción de Videojuegos Aplicados a Salud (IPVÍAs)”, financiado por la Secretaría de Políticas Universitarias (SPU) según Resolución SPU-2919-2013. También a través del proyecto “Videojuegos para motivar conductas Saludables” presentado y aprobado en el marco de la Convocatoria 2014 – 2015.

El Proyecto MoVer-T se llevará a cabo en el Instituto de Informática de la Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de San Juan. Es una investigación aplicada, en la que se ha priorizado una metodología cuanti-cualitativa y se ha optado por un diseño descriptivo y experimental. Las experimentaciones se realizarán con usuarios reales, en este caso, pacientes en proceso de rehabilitación de la Fundación Abril (institución destinataria). La ejecución del proyecto está planificada en las siguientes fases: Fase I: Modelización de los procesos de interacción, Fase II: Desarrollo de los productos de software, Fase III: Implantación de los productos de software, que se describen a continuación en la introducción.

Introducción

La parálisis cerebral (PC) es un tipo de trastorno que afecta tanto a niños como a adultos, con una prevalencia de dos por cada 1000 nacidos vivos [6]. La esperanza de vida de personas adultas con PC es similar al de la población en general. Esta situación plantea nuevos desafíos en el ámbito de la neurorehabilitación[7]. Si bien los programas de neurorehabilitación actuales se centran en la terapia convencional y práctica de ejercicios motrices [2]; existen evidencias de que aplicaciones de Realidad Virtual (RV) orientadas a

rehabilitación de pacientes con PC han provocado un aumento en el interés, el apego y la motivación por la terapia[1]. Esto lleva a suponer, por lo tanto, que los videojuegos pueden favorecer la rehabilitación de personas con discapacidades motrices. En este sentido, expertos en rehabilitación motriz de personas con hemiplegia o parálisis cerebral [5] sostienen que la ejercitación llevada a cabo mediante videojuegos de realidad virtual[4] en la propia casa de los pacientes constituye el futuro de las terapias de rehabilitación[3]. Debido a su naturaleza inmersiva, la realidad virtual (VR) está siendo muy utilizada en este tipo de terapias, por su capacidad de representar situaciones de la vida real en un contexto de total seguridad[8]. En los proyectos predecesores se utilizó la tecnología de Google Cardboard [9] y Unity3D[10], por sus características y facilidad de aplicación. En el presente proyecto, se analizarán y utilizarán otras tecnologías, como plataforma de Samsung GearVR [11], que poseen algunas características muy adecuadas para facilitar la interacción del paciente con un mundo virtual.

El Proyecto MoVer-T: es un proyecto de I+D que se sustenta en las siguientes hipótesis:

H1) Los pacientes que hagan uso de la herramienta de realidad virtual en forma periódica, presentarán una mejora en los indicadores de progreso de su rehabilitación.

H2) Los profesionales de la salud que adopten esta tecnología, podrán realizar un seguimiento más preciso de sus pacientes en forma continua y distribuida, potenciando así el alcance de su profesión.

H3) La interfaz de realidad virtual en línea facilitará a los pacientes, su integración en comunidades de jugadores en línea, motivando su crecimiento

cognitivo y social. Se encuadra en el marco de la investigación aplicada, en la que se ha priorizado una metodología cuanti-cualitativa. Se ha optado por un diseño descriptivo y experimental. Las experimentaciones se realizarán con usuarios reales (pacientes en proceso de rehabilitación de la Fundación Abril) para la evaluación y validación del entorno de monitoreo remoto y del videojuego en realidad virtual desarrollado. El relevamiento de información se llevará a cabo a través de: Entrevistas focalizadas a informantes claves seleccionados: docentes y padres de la Fundación Abril; Instancias de grupos focales (focus groups) con profesionales de la salud de las distintas disciplinas de la Institución destinataria y de otras de características similares que atienden la problemática de los sujetos bajo estudio. El criterio de selección de los mismos en el marco de esta investigación son los pacientes cuya edad cognitiva se encuentra comprendida entre 5 a 10 años. También se aplicará la Observación con participación moderada realizada por investigadores y desarrolladores, durante las experimentaciones con los prototipos de videojuegos en realidad virtual.

Los documentos generados y el desarrollo del software se realizarán teniendo en cuenta 3 fases: la Fase de Preproducción, en la que se realizará prototipado, prueba, retroalimentación y prototipado segunda versión, y posterior generación del Documento de diseño del videojuego (GDD). La Fase de Producción, en la que se crean los modelos, música y/o gráficas del videojuego y se genera y prueba el primer nivel del juego. Y por último la Fase de Postproducción, donde se llevarán a cabo las Pruebas de stress del videojuego en diversos hardware y plataformas y se corrigen de bugs y mejoras del software.

Líneas de Investigación, Desarrollo e Innovación

Este proyecto se encuadra dentro de los procesos de Ingeniería de Software aplicados a videojuegos para la salud; y dentro del Diseño de Videojuegos para personas con capacidades especiales.

Resultados y Objetivos

Los resultados alcanzados en el marco de los proyectos predecesores son la generación de conocimiento sobre el diseño y desarrollo de videojuegos para realidad virtual. En donde se investigaron tecnologías de desarrollo y prototipación de videojuegos. Luego, a partir de este conocimiento se desarrollaron prototipos de videojuegos destinados a la rehabilitación motriz de pacientes con parálisis cerebral y se realizaron pruebas experimentales con algunos usuarios para poder mejorar los prototipos.

Uno de tales desarrollos se plasmó en la tesis de trabajo final de la Licenciada María Pelozo denominada “Definición y Experimentación de Metodología de Preproducción Basada en la Prototipación para el Desarrollo de Videojuegos para la Salud”. En dicha tesis, se experimentó con un entorno de realidad virtual ambientado en una historia gráfica, en la que el niño debe encontrar ciertos elementos en un laberinto siguiendo pistas. Se pretende seguir avanzando en la generación de conocimiento sobre el desarrollo de videojuegos para realidad virtual orientados a la rehabilitación motriz, mediante la interacción permanente con un equipo multidisciplinario de expertos tanto en los aspectos tecnológicos como en terapias de rehabilitación motriz de pacientes con parálisis cerebral para luego desarrollar videojuegos en entornos telecontrolados.

El Proyecto MoVer- T tiene como objetivo desarrollar un videojuego de realidad virtual multijugador en línea (Multiplayer Online Virtual Reality Game, o MOVR-Game) con dos aplicaciones interconectadas: una para los pacientes que se ejecutará en los lentes de realidad virtual; y otra para el profesional de la salud, quien desde su consultorio podrá realizar la gestión, seguimiento y adaptación de las sesiones de juego de cada uno de sus pacientes. La tecnología desarrollada tendrá como destinatarios a los pacientes de la Fundación Abril, que brinda rehabilitación a personas con discapacidad entre otras actividades. Se espera que los datos obtenidos durante las experimentaciones contribuyan a mejorar la calidad de las terapias de rehabilitación asistidas por computador mediante el desarrollo de videojuegos adaptados a las necesidades de los pacientes con parálisis cerebral y éstos no se difundirán a no ser que sea autorizado por la institución beneficiaria y/o por los tutores/responsables de los usuarios involucrados. De mediar esta autorización los datos no se vincularán con la identidad real de los usuarios participantes, utilizándose para tal fin apodosos o nombres ficticios.

Formación de Recursos Humanos

El equipo está conformado por los siguientes miembros:

Ormeño, Emilio Gustavo. Director del proyecto. Actualmente doctorando del Doctorado en Ciencias de la Informática, y cuya tesis está orientada al área de Videojuegos para la salud.

González, Mónica Gilda. Co-directora. Actualmente desarrollando su tesis Doctoral en Educación, en instancia de Informe Final, cuyo tema refiere a procesos de apropiación de nuevas tecnologías en Adultos Mayores. En el

marco del Doctorado en Educación de la Universidad Católica de Cuyo

Zapata, Sergio Gustavo. Integrante del proyecto. Actualmente doctorando del Doctorado en Ciencias de la Informática, y cuya tesis está orientada al área de Ingeniería de Software Global.

Ferrarini Oliver, Cintia. Integrante del proyecto. Actualmente doctorando del Doctorado en Ingeniería de la Universidad Nacional de Cuyo. Cuya área de investigación son los Procesos Estocásticos de Markov para la planificación automática de servicios. Magister en Informática de la UnLaM.

Olguín, Luis Alberto. Integrante del proyecto. Actualmente maestrando de la Maestría en Informática de la Universidad Nacional de San Juan. Especializado en enseñanza de programación de videojuegos para niños y adolescentes con plataformas como Scratch y App Inventor.

Trabajos de Tesis: 1) Pelozo, María Lidia. “Definición y Experimentación de Metodología de Preproducción Basada en la Prototipación para el Desarrollo de Videojuegos para la Salud”. Asesores: Sergio Zapata - Emilio Ormeño. Año 2015. **2)** Belchior Florencia. “Sistema Web con Gamificación (SiWGa) para Promover Competencias de Resolución de Problemas Matemáticos en Nivel Primario”. Asesoras: Mg. Mónica González - Mg. Cintia Ferrarini. Por presentar en abril de 2016. **3)** Avendaño, Mauro. “Plataforma para videojuegos de realidad virtual multijugador en línea”. Asesores: Mg. Emilio Ormeño, Mg. Cintia Ferrarini. Por presentar en junio de 2016. **4)** Benzonelli, Juan José. (Tesis de Maestría en Informática). “Técnica de Gamificación para mitigar aspectos insalubres de sistemas de información”. “A Gamification Technique for mitigating unhealthy issues of

information systems”. Asesor: Mg. Emilio Ormeño. 5) González, Gemma. “Proceso Distribuido de desarrollo de Videojuegos para la Salud”. Asesor: Mg. Sergio Zapata.

Becario: Rodrigo Kokot. “Captura de parámetros corporales utilizando Xbox 360 con Kinect”. Asesor: Mg. Emilio Ormeño.

Referencias

[1] Chen YP (2007). Use of Virtual Reality to improve upper extremity control in children with Cerebral Palsy: A single subject design. *Physical Therapy* (2007); 87(11):1441-57

[2] Duff S. V, Gordon AM (2003). Learning of grasp control in children with hemiplegic cerebral palsy, *Dev Med Child Neurol* 45, 746-757

[3] Golomb M. R., Barkat-Masih M., Rabin B., AbdelBaky M., Huber M., & Burdea G.(2009), “Eleven Months of home virtual reality telerehabilitation - Lessons learned,” in *Virtual Rehabilitation International Conference*, 2009, pp. 23–28.

[4] Leston J.(1996), “Virtual reality: the IT perspective,” *Comput. Bull.*, vol. 38, no. 3, pp. 12–13, Jun.

[5] Raina P., O’Donnell M., Rosenbaum P., Brehaut J., Walter S. D., Russell D., Swinton M., Zhu B., & Wood E., (2005) “The health and well-being of caregivers of children with cerebral palsy,” *Pediatrics*, vol. 115, no. 6, pp. e626–e636.

[6] Rosenbaum P, Paneth N., Leviton A, et. Al (2007).A report: the definition and

classification of cerebral palsy April 2006. *Dev Med Child Neurol Suppl*; 109:8-14

[7] Tosi L, Maher N, Winsolr-Moore D, Murray DO, Laisen M (2009). Adults with cerebral Paisy: A workshop to define the challenges of treating and preventing secondary musculoskeletal and neuromuscular complications in this rapidly growing population. *Mac Keit Press Developmental Medicine & Child Neurology*, 51 (Suppl. 4):2-11

[8] Ustinova K. I., Ingersoll C. D., & Cassavaugh N.(2011). “Short-term practice with customized 3D immersive videogame improves arm-postural coordination in patients with TBI,” in *2011 International Conference on Virtual Rehabilitation (ICVR)*, pp. 1–7.

[9] “Google Cardboard,” *Wikipedia*, the free encyclopedia. 11-Mar-2015.

[10] “Unity - Game Engine.” [Online]. Available: <http://unity3d.com/es/>. [Accessed: 15-Dec-2015].

[11] “Samsung GALAXY Note Gear VR - Descripción.” [Online]. Available: <http://www.samsung.com/latin/promotions/feature/gearvr/>. [Accessed: 15-Dec-2015].

Mejoras en la ejecución de BPM incluyendo conceptos de Green IT

Javier Díaz¹, Patricia Bazán¹, Anahí S. Rodríguez¹, Viviana¹ M. Ambrosi²

1- LINTI – Facultad de Informática – UNLP

2- – Comisión de Investigaciones Científicas de la provincia de Buenos Aires –
CIC-BA

javier.diaz@info.unlp.edu.ar, pbaz@info.unlp.edu.ar, anahi@linti.unlp.edu.ar,
vambrosi@info.unlp.edu.ar

Resumen

Hoy en día las organizaciones y los gobiernos están más interesados y comprometidos con la ecología. Surgen nuevos retos que conllevan a redefinir las formas de trabajo para minimizar impactos en el medio ambiente. Los procesos de negocio no están ajenos a esta realidad, por lo que las organizaciones deben trabajar teniendo en cuenta esta problemática inminente.

En esta investigación se analizan conceptos asociados a Green IT y Green BPM y la manera de combinarlos (usando Social BPM) para aportar una mejora en la ejecución de los procesos. Se plantea una monitorización dinámica y con rastros enriquecidos que aporten a la mejora continua de un proceso "verde".

Palabras clave: Desarrollo sostenible – Green IT – Green BPM — Social BPM – Monitorización

Contexto

Las investigaciones realizadas dentro del grupo de trabajo del LINTI (Laboratorio de Investigación en Nuevas Tecnologías) [1] en torno a Green IT y

BPM, se vinculan al sector productivo y su aplicación en el ámbito social. Se pretende enfatizar con la presente propuesta dicha perspectiva práctica e innovadora del uso de la tecnología, enlazando dos temáticas que atañen a las organizaciones actuales, tanto públicas como privadas, que persiguen objetivos que mejoren su productividad pero sin descuidar la sostenibilidad de sus acciones.

Introducción

Hoy en día las organizaciones y los gobiernos están más interesados y comprometidos con la ecología. Surgen nuevos retos que conllevan a redefinir las formas de trabajo para minimizar impactos en el medio ambiente. Así las organizaciones y el sector de TI (Tecnología de la Información) juegan un papel muy importante para implementar o verificar métodos, políticas o procedimientos que ayuden a la conservación del medioambiente y realicen sus tareas de manera responsable [2].

Existen muchas definiciones para Green IT, como por ejemplo Jonh Lamb en [3] define Green IT como "El estudio y la

práctica de la utilización de los recursos informáticos de manera eficiente" y Weber y Wallace en [4] la definen como "La reducción del impacto ambiental del funcionamiento del departamento de TI". J. Díaz et al lo definen como "uso eficiente de los recursos computacionales minimizando el impacto ambiental, maximizando su viabilidad económica y propiciando una mejora del contexto social" [5].

De estas definiciones se desprende que Green IT trata de considerar la protección del medio ambiente todo el ciclo de vida de TI. Existen algunas prácticas como la eficiencia en los servidores de datos o la gestión de residuos electrónicos, siendo éstas sólo algunas de las posibles acciones de Green-IT [6]. Dichas prácticas se pueden evaluar con métricas para hardware y software que determinen el impacto negativo y permitan realizar acciones para revertir o mejorar las actividades o procesos en las organizaciones [7], [2].

Los procesos de negocio no están ajenos a esta realidad, por lo que las organizaciones deben trabajar teniendo en cuenta esta problemática inminente.

Las organizaciones que adhieren a un paradigma de gestión por procesos y adhieren a BPM, se introducen en un ciclo para identificar, modelar, desplegar y monitorizar procesos de negocio en búsqueda de la mejora continua. Este ciclo se cierra con la monitorización de indicadores que dan cuenta del rendimiento de los procesos y que suelen definirse de manera clásica y universal, como por ejemplo, tiempo de ejecución entre ejecución de actividades de un proceso, cantidad de tareas detenidas, en progreso o fuera de término. En este sentido, sería razonable pensar que estas mismas organizaciones comprometidas con el medio ambiente, estarían interesadas en medir cuan sostenibles son

sus procesos, por ejemplo, en función de los recursos que consumen.

Surge así el concepto de proceso "verde" que requiere que se definan indicadores de mejoras vinculada a aspectos ecológicos.

Estado del arte en Green IT aplicado a BPM

Green BPM se encuentra en las primeras fases de investigación, es la suma de todas las actividades de gestión soportadas por los sistemas de información que ayudan a monitorizar y reducir el impacto ambiental de los procesos de negocios en sus etapas de diseño, mejora, implementación u operación, así como guiar hacia un cambio cultural dentro del ciclo de vida del proceso [8]. La bibliografía es muy reciente en cuanto a cómo medir y monitorizar un Green BPM [9]. Se encuentran definidos modelos de madurez de procesos de negocio en [10] los cuales miden 6 criterios: cultura, métodos, tecnología, recursos humanos, lineamientos estratégicos y gobernabilidad. Para alcanzar el objetivo de procesos "verdes", podemos agregar a estos criterios conceptos de Green IS y Green IT [9].

Para abordar un Green BPM, podemos tener en cuenta los siguientes puntos [9]:

- Actitud Verde: describe la actitud de la empresa y los empleados para obtener resultados sostenibles, viendo la responsabilidad y cuan interesados están en la adopción de medidas de mejora para lograr un proceso más sostenible. Este es el factor fundamental para tener un Green BPM a largo plazo. Tomando conciencia del impacto ecológico negativo del proceso de negocio haciéndolo más sostenible para tener un impacto menos negativo.
- Estrategia Verde: la gestión del Green BPM debe estar relacionada con la

estrategia comercial para el ciclo de vida del BPM, para poder establecer mejoras ecológicas. Se planifica la estrategia y así genera un marco de actuación sostenible. Esto es importante para la adquisición de una estrategia holística para todo el ciclo de vida.

- **Gobernabilidad Verde:** esta etapa junto con la estratégica son parte de las tareas de gestión. Define la estructura de la gestión, los roles, autoridad para toma de decisiones, responsabilidades y etc. y el control de acciones verdes. También se organizan las personas involucradas en el proceso. La asignación de presupuestos y otros recursos son importantes para Green BPM. Por otra parte, la gobernabilidad verde es responsable de la definición de las metas y de medir el progreso hacia ellos.

- **Modelado verde:** se considera el impacto ambiental en el modelo del proceso. Se establecen valores y factores claves para cada unidad de la empresa. Con estos valores podemos medir valores claves ecológicos en los procesos de negocio, sub-procesos, actividades, recursos. Pudiendo asignar valores a cada proceso de los recursos utilizados, la producción de residuos, o emisión de CO².

- **Monitoreo verde:** es de gran utilidad tener medidas para monitorear el impacto ambiental. El monitoreo verde utiliza indicadores ecológicos clave de rendimiento (KPI - *Key Performance Indicators*) para cada actividad de los procesos de negocio. Con estos valores se pueden medir su impacto ambiental. Los KPI pueden ser de ayuda a una organización para cuantificar las TIC y evaluar el desempeño en esta área [16]. Los indicadores claves ecológicos (KEI) son de tipo estáticos y dinámicos. El dinámico es el que se genera del uso real de un dispositivo, por ejemplo el consumo de energía de una impresora;

mientras que los estáticos son independientes del uso, por ejemplo servidores, en el cual su uso es constante. Para el caso de los KEI dinámicos se utilizan redes de sensores para medir automáticamente. El monitoreo verde mide los niveles de KEIs de la organización, teniendo en cuenta valores internos y externos.

- **Optimización verde:** luego de obtener las mediciones anteriormente citadas se deberá realizar un análisis para evaluar posibles cambios en el proceso de negocio [13], si es que es necesario, optimizando con estas mejoras el impacto climático. Como primera medida no es necesario el cambio de todo el proceso, sino por etapas, por ejemplo utilizando más formato digital que documentos impresos.

Indicadores de aspectos ecológicos en BPM.

En [11] se propone como extender el modelado clásico de BPM agregando notaciones que indican aspectos ecológicos, como por ejemplo, agregando una notación para las actividades que generan emisiones de gases las cuales hacen uso de combustible, actividades que utilizan papel, indicadores para mostrar la cantidad de gases que se emiten en el pool o lane, etc. Estos indicadores de aspectos ecológicos podrían dar un gran valor agregado a los resultados esperados del proceso de negocio. En el cual el monitoreo del mismo tendría que tener en cuenta estos rastros de ejecución.

Rastros enriquecidos a través de Social BPM

Para aportar dinamismo a la ejecución del proceso de negocio en [12] se especificaron características funcionales que debe tener un motor de ejecución de procesos para incorporar la interacción

entre participantes y enriquecer los rastros de ejecución de sus actividades, como por ejemplo, incorporación de notificaciones, comentarios etiquetados, definición de tareas ad-hoc sin modelar y notificaciones a los responsables de los procesos.

En [12] se presenta una mejora de ejecución de procesos incluyendo aspectos de distribución y enriquecimiento de rastros para obtener una monitorización más efectiva e informada aportando al ciclo de mejora continua.

El dinamismo brindado por la mejora propuesta en [12] puede ser utilizado para incluir en "tiempo real" los valores de diversos aspectos ecológicos (por ejemplo: emisión de gases, consumo de electricidad, de papel, etc). Dichos valores se incluirán como "comentarios etiquetados", por ejemplo, asociados a las actividades pertinentes para luego tener una estimación global del impacto ambiental producido por la ejecución del proceso. Con este enfoque los rastros de ejecución ya no serán los clásicos sino que son propios de la ejecución del proceso. Los datos asociados a cada actividad pueden obtenerse en tiempo real o no, dependiendo de la naturaleza de la actividad. Esto nos lleva a cambiar el enfoque de monitorización estática a una dinámica.

A través de la monitorización dinámica se pueden emitir alertas que sirvan para analizar el uso de las mejoras propuestas en los patrones definidos en [13] para obtener un Green BPM partiendo de un BPM clásico. Por ejemplo, aunque en el modelado del proceso inicial no se haya tenido en cuenta el indicador de consumo de energía, si se detecta dinámicamente que un servidor de la organización tiene un consumo de energía excesivo, se

puede aplicar el "Patrón 3: Cambio de Recursos" el cual propone reemplazar el servidor propio por un servidor en la nube, o un patrón más adecuado en la planificación de la adquisición o en la generación de residuos electrónicos.

Líneas de Investigación, Desarrollo e Innovación

La incorporación de aspectos de Green IT en la ejecución del proceso de negocio conlleva a plantear nuevos y diferentes criterios para monitorizar dichos procesos.

En este nuevo escenario, la monitorización debe considerar la medición de rastros enriquecidos en tiempo real, que pueden haber sido definidos o no dentro de la etapa de modelado.

Para esto se toma como base la línea de trabajo vinculada a las mejoras en la monitorización de procesos no clásicos que propone el enfoque de BPM Social.

En este sentido, lo que se denomina "socialización de procesos de negocio" aportan un enfoque para medir y evaluar procesos a partir de datos enriquecidos, cómo lograr el monitoreo dinámico con esta nueva información [12].

Resultados y Objetivos

El objetivo de esta investigación es plantear un mecanismo para incorporar indicadores en los procesos de negocio que permitan establecer nuevas medidas de rendimiento en tiempo real y que logren monitorizar los procesos de negocio de acuerdo a parámetros de sostenibilidad y no sólo de productividad. Asimismo, se pueden obtener tableros de comando que las organizaciones utilicen para la toma de decisiones, donde las variables a considerar puedan provenir

del exterior de manera dinámica y en tiempo real.

Formación de Recursos Humanos

BPM y la mejora continua de procesos de negocio aplicada a los ambientes de ejecución, es una línea de trabajo que ha formado docentes e investigadores entorno a la solución de problemas reales. Paralelamente, los aspectos de Green IT también se vienen trabajando e incluso se han insertado como temática en las carreras de grado de la Facultad de Informática, acompañando por un trabajo de extensión universitaria sobre los residuos electrónicos desde el programa E-Basura [14] de la Universidad Nacional de la Plata [15].

Cada una de estas líneas de trabajo cuenta hoy con formación en recursos humanos en torno a la definición de tesis de grado, dejando así líneas de trabajo abiertas para continuar con las investigaciones, vinculando ambos enfoques.

Referencias

- [1] <http://www.linti.unlp.edu.ar/>
- [2] Javier Muñoz Giner, Yuresky Rojas Rincón, “Nuevas tendencias en tecnologías verdes -Green IT para la Gestión en Organizaciones”, II Congreso Iberoamericano SOCOTE - Soporte al Conocimiento con la Tecnología- y VII Congreso SOCOTE Universidad Politécnica de Valencia, Noviembre 2010
- [3] Jonh Lamb, “The greening of IT: how companies can make a difference for the environment”, Capítulo 1. IBM Press – 0137150830/9780137150830 – Mayo 2009
- [4] Lawrence Webber, Michael Wallace, “Green Tech: How to Plan and Implement Sustainable IT Solutions”, Julio 2009.
- [5] Javier Díaz, Viviana Ambrosi, Néstor Castro, Claudia Banchoff Tzancoff, Marcelo Raimundo, “Porque incluir Green IT en la currícula de Informática”
- [6] Fatima Zahra HANNE, “GREEN-IT: Why Developing Countries Should Care?”, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, Julio 2011.
- [7] <http://green-software-engineering.de/en/publications.html>
- [8] Nicky Opitz, Henning Krüp, Lutz M. Kolbe, “Green Business Process Management – A Definition and Research Framework”, 47th Hawaii International Conference on System Science, 2014
- [9] Nicky Opitz, Henning Krüp, Lutz Maria Kolbe, “Environmentally sustainable business process management – developing a green bpm readiness model”.
- [10] Maximilian Röglinger, Jens Pöppelbuß, Jörg Becker, “Maturity Models in Business Process Management”, “Business Process Management Journal 18” (2012).
- [11] Jan Recker, “Green, Greener, BPM?”, <http://www.bptrends.com/publicationfiles/07-05-2011-COL-Class%20Notes--Green%20Greener%20BPM-Recker.pdf>.
- [12] Patricia Bazán, Roxana Giandini, Jose Martinez Garro, Javier Diaz, “Mejora de la monitorización y ejecución de procesos de negocio con integración y socialización”, Latin American Computing Conference (CLEI), 2015
- [13] Alexander Nowak, Frank Leymann, Daniel Schleicher, David Schumm, Sebastian Wagner, “Green Business Process Patterns”, Proceedings of the 18th Conference on Pattern Languages of Programs, 2011
- [14] <http://e-basura.linti.unlp.edu.ar/>
- [15] <http://www.unlp.edu.ar/>
- [16] ITU General specifications and KPIs https://www.itu.int/dms_pub/itu-t/oth/4B/04/T4B0400000B0009PDFE.pdf

ESTUDIO DE ADOPCIÓN DE TÉCNICAS DE DESARROLLO DE SOFTWARE GUIADO POR LAS PRUEBAS

Calixto Maldonado, Iris Gastañaga, Claudia Inchaurredo, Pablo Vaca, Matías Bueno, María Soledad Romero, Juan Pablo Peretti

Grupo de Investigación, Desarrollo y Transferencia de Sistemas de Información (GIDTSI)/ Departamento Ing. Sistemas de Información / Facultad Regional Córdoba / Universidad Tecnológica Nacional

Maestro López y Cruz Roja

calixtomaldonado@hotmail.com, vacapablo72@gmail.com, matiasbueno@gmail.com, cinchaurredo@hotmail.com, romeroma.soledad@gmail.com, peretti.juan@gmail.com

Resumen

En el presente proyecto, dentro del Grupo de Investigación, Desarrollo y Transferencia de Sistemas de Información (GIDTSI), se estudiaron las implementaciones existentes de los frameworks aplicados en Test-Driven Development (TDD) y Behavior-Driven Development (BDD) y como resultado al final del mismo se prevé proponer una forma de integración entre distintas técnicas de (TDD) y (BDD). Para el estudio de campo entre las empresas de desarrollo de software se ha creado un formulario de encuesta con fecha próxima de publicación para ser completado por responsables de área de Desarrollo de software y de Testing.

Palabras clave: TDD, BDD, Ingeniería de Software, testing

Contexto

En la Facultad Regional Córdoba de la Universidad Tecnológica Nacional el Grupo GIDTSI tiene como una de sus líneas de investigación a la Ingeniería en Sistemas de Información y de Software, incluyendo proyectos de investigación en

calidad, métricas, mantenimiento, trazabilidad, requerimientos, bases de datos, seguridad, relacionados con las áreas de conocimiento: Sistemas de Información, Gestión Ingenieril y Programación.

A inicios del 2013 se observó en el país una fuerte tendencia a aplicar esfuerzos en TDD (Test-Driven Development) y además, se estaba empezando a trabajar en BDD (Behavior-Driven Development). Se conocía que para tal fin existían diferentes frameworks tales como JUnit o SUnit, los cuales se encontraban en un avanzado grado de desarrollo e implementación.

Si bien todas las metodologías de desarrollo de software basadas en las pruebas son una disciplina relativamente nueva, ya que han cobrado auge en los últimos 10 años, podemos nombrar algunos referentes a nivel internacional como Kent Beck (1) y Martín Fowler (2). En el ámbito local, actualmente en la UTN Facultad Regional Córdoba no hay grupos trabajando en la materia, y en el resto del país solamente un grupo mixto

del LIFIA (UNLP) y de la Facultad de Ingeniería de la UBA están investigando la aplicación de estos enfoques al desarrollo de sistemas web.

Introducción

El desarrollo de software se ha transformado a lo largo del tiempo de un arte a un proceso detallado tendiente a obtener software de alta calidad, mantenible y por sobre todo confiable a lo largo del tiempo (1) (2). Es conocido en el ámbito del desarrollo de software que el 70% del costo de un software a lo largo de la vida útil del mismo es consumido en la etapa de mantenimiento (3) (4), por lo tanto es necesario encontrar mecanismos, ya sean manuales o automáticos que nos permitan garantizar que lo que se encuentra funcionando en un software siga funcionando luego de la introducción de nuevas características.

El desarrollo de software guiado por el negocio (BDD y STDD por sus siglas en inglés) o el desarrollo de software guiado por las pruebas unitarias o de integración (TDD y ATDD por sus siglas en inglés y todas sus variantes) son dos alternativas que buscan garantizar que el desarrollo progresivo e incremental del software se hace de manera controlada y salvaguardando el software ya desarrollado, son dos técnicas que se pueden implementar en conjunto o por separado, en ambos casos son técnicas que permiten automatización logrando de esta manera la reducción en los tiempos de pruebas unitarias, pruebas de regresión y pruebas de integración (5).

Sin embargo, varios investigadores y practicantes (5) han alertado que la proliferación de enfoques al desarrollo de software basado en pruebas disminuye su aplicabilidad en la industria. Es por esto que durante el proyecto de investigación se propuso desarrollar las siguientes actividades:

- Investigar los procesos de desarrollos basados en la generación de las pruebas como primer paso en el desarrollo de software (BDD, TDD, ATDD, STDD, NDD, DDD) comparando los diversos enfoques según la óptica de diversos autores y agregando nuestro punto de vista basado en la experiencia profesional en el ámbito privado utilizando estas metodologías.

- Investigar los distintos software y arquitecturas existentes para la automatización de las pruebas, comparando los mismos, instalando y probando cada uno de ellos, generando conclusiones y exponiendo las mismas en documentación con el objetivo de compartir con la comunidad académica nuestras conclusiones. A modo de ejemplo de los software a investigar se citan los siguiente: Jbehave (6), Selenium (7), JUnit (8) (9), NUnit, easyb (10) (11), RSpec(12), HttpUnit. Las pruebas se ejecutaro sobre dos plataformas de sistemas operativos una en Linux y otra sobre Windows, y como lenguajes de programación se utilizarán Java, Visual C++, Visual Basic, HTML, Javascript, en el caso de servidores web se utilizará Apache Tomcat.

- Investigar los software y arquitecturas asociadas al proceso, el cual se utiliza para medir la calidad de las pruebas y la cobertura que tienen los test sobre el código que se está probando, tales como Maven(14) (15), Visual Studio (16) (17), SimpleCov(18) (19). Se instaló el software y se ejecutaron pruebas del mismo.

- Como resultado de la investigación se generará una propuesta de integración de varios enfoques y se estructurará una propuesta de desarrollo de un framework (software y arquitectura) que permita la integración automática, facilitando la aplicación de estos métodos de desarrollo en el ambiente industrial.

- A modo de transferencia de los conocimientos adquiridos y generados en los pasos anteriores, se trabajará en la difusión de los resultados en cuatro aspectos:

1. Se buscará interesar a las cátedras de análisis, diseño y programación para realizar intercambio de transferencia al aula de los avances logrados en este trabajo.

2. Se buscará interesar a los laboratorios de desarrollo de software existentes en la universidad para el intercambio de experiencias y realizar transferencia de los avances en el proyecto.

3. Se enviaron artículos de difusión y científicos a los siguientes congresos WICC, JAIIO.[1] [2].

Líneas de Investigación, Desarrollo e Innovación

La investigación completa se basó en procesos de investigación y se utilizarán técnicas deductivas y exploratorias, pero siempre apuntando a una investigación aplicada.

Uno de los objetivos del presente proyecto es realizar los contactos con grupos que se encuentren trabajando en la materia, tanto a nivel nacional como en el extranjero, con el fin de crear vínculos y lograr intercambio de conocimientos. Por otra parte, esta metodología está cobrando importancia dentro de las empresas que desarrollan software y se está volviendo una práctica normal y obligatoria durante el ciclo de vida de los proyectos de software. Por ejemplo, la empresa en la cual trabaja el Ing Pablo Vaca (Globant) es una de las que está dando mucho impulso al uso de estas prácticas.

Actualmente los más utilizados son los basados en BDD y TDD. Integrantes del equipo de investigación trabajan habitualmente en el ámbito privado con ambas técnicas. Se disponen de conocimientos sobre ambas técnicas pero con implementaciones separadas, mediante el presente proyecto se buscará profundizar en la integración entre distintas técnicas.

Por otra parte los resultados obtenidos del proyecto son la base para la formación de recursos humanos capacitados para desenvolverse en el ámbito privado y público, así como la aplicación de la gama de servicio que ofrece el laboratorio

de desarrollo de Software, generando de esta forma puestos de trabajo de alta nivel y requeridos a nivel mundial.

Las actividades fueron: una etapa exploratoria, en la cual se hizo un estudio bibliográfico de fuentes conocidas e investigar nuevos trabajos publicados sobre la materia, a la cual le siguió una etapa de investigación aplicada, probando productos de código libre y abierto disponible y comparando sus funcionalidades.

Para la elaboración de papers se trabajó planteando hipótesis y luego mediante investigación bibliográfica, intercambio con otros grupos y pruebas desarrolladas sobre el software instalado se buscará demostrar la misma o explicar por que dicha hipótesis no es correcta.

Para la instalación de software y entornos de ejecución de pruebas se utilizaron técnicas deductivas, partiendo de la lectura de la documentación de cada software y de foros que traten el tema, documentado cada paso y situación a resolver, con la finalidad de armar una base de conocimiento sobre el tema, que luego sirva para la elaboración de documentación de base en las transferencias que se realicen.

Las tareas de formación de recursos humanos se basaron en un plan específico de formación, previendo se realizar al menos dos presentaciones abiertas a la comunidad en general y con invitación a las cátedras de programación e Ingeniería de Software de la Carrera de Ingeniería en Sistemas de Información de la UTN-FRC

de los temas principales del proyecto. Se prevé publicar en la web la presentación electrónica, y si los medios disponibles lo permiten obtener un registro en video de la misma.

Objetivos Generales del Proyecto:

- Investigar la posible integración de las distintas metodologías de desarrollo guiado por pruebas.
- Proponer un marco teórico para la integración de algunas de estas metodologías y establecer los requerimientos de un framework que permita automatizar la aplicación integrada de dichas metodologías.

Objetivos Particulares del Proyecto:

- Estudiar las metodologías de desarrollo guiado por pruebas y establecer un foco de transferencia de resultados hacia la universidad y la industria.
- Instalar un entorno real de pruebas que le permitan al equipo experimentar y poner en práctica las conclusiones alcanzadas durante la investigación.
- Formar recursos humanos especializados en desarrollo de software guiado por pruebas.
- Realizar talleres de capacitación en los laboratorios CIDS y LIDICaSo de la FRC
- Asistir técnicamente en las mejores prácticas a grupos de Investigación de la

disciplina Homologados en la SCyT de la FRC – UTN Exponer los resultados alcanzados y los objetivos en curso/futuros de la línea de I/D/I presentada.

Formación de Recursos Humanos

Un punto importante es la formación de profesionales capacitados en el desarrollo de software guiado por las pruebas.

Esta formación debe incluir desde los aspectos teóricos hasta los de implementación práctica. Pero también se busca que los integrantes, becarios y tesistas adquieran experiencia en la investigación y en el futuro puedan replicar y llevar adelante otros proyectos de investigación.

La experiencia adquirida en este proyecto es aplicada dentro de la universidad en grupos de investigación, cátedras de grado y en el ámbito privado.

Con estas habilidades se propiciaron que los participantes accedan a la categoría inicial de investigador dentro de la docencia investigativa.

Otro aporte a la formación de recursos humanos es la elaboración de tesis de maestría de integrantes del proyecto, con lo cual se podrá alcanzar mayores niveles académicos que permitan el progreso en el ámbito docente y profesional.

Referencias

- (1) - Ingeniería de Software - Somerville - Addison Wesley.
- (2) - Ingeniería de Software - Pfleeger - Prentice Hall.

(3) - SW Maintenance, Concepts and Practices - Takang, Grubb - World Scientific 2005

(4) - SW Engineering Body of Knowledge (IEEE 2004 version).

(5) - Estado del arte y tendencias Test-Driven Development - Moisés, Carlos Fontela - 2011 - Trabajo Final Integrador de la Especialización en Ingeniería de Software - Universidad de La Plata - Facultad de Informática -

(6) - <http://jbehave.org/>

(7) - <http://seleniumhq.org/>

(8) - <http://www.junit.org/>

(9) - <http://es.wikipedia.org/wiki/JUnit>

(10) - <http://code.google.com/p/easyb/>

(11) - <http://www.ibm.com/developerworks/java/tutorial/s/j-easyb/>

(12) - <http://rspec.info/>

(13) - <http://en.wikipedia.org/wiki/RSpec>

(14) - <http://maven.apache.org/>

(15) - <http://en.wikipedia.org/wiki/Maven>

(16) - http://es.wikipedia.org/wiki/Microsoft_Visual_Studio

(17) - <http://msdn.microsoft.com/es-ar/vstudio/aa718325.aspx>

(18) - <http://rubygems.org/gems/simplecov>

(19)- <http://www.becodemyfriend.com/tag/simplecov/>
Una exposición general sobre el tema de I/D/I con las referencias bibliográficas relacionadas. Se sugiere no más de 3 páginas.

Citas de los autores

[1] Vaca, P. A., Maldonado, C., Inchaurredo, C., Peretti, J., Romero, M. S., & Bueno, M. (2013). Test-Driven Development-Una aproximación para entender su utilidad en el proceso de desarrollo de Software. *Conaiisi-Cordoba, Octubre de*, 2347-0372.

[2] Vaca, P. A., Maldonado, C., Inchaurredo, C., Peretti, J., Romero, M. S., Bueno, M., & Cagliolo, M. Test-Driven Development-Beneficios y Desafíos para el Desarrollo de Software.

Mejora y gestión de procesos de negocio inter-organizacionales aplicando técnicas de minería de procesos

Lazarte, Ivanna¹, Acosta Parra, Carlos¹; Vilallonga, Gabriel^{1,2}

¹Departamento Sistemas/Facultad de Tecnología y Ciencias Aplicadas/Universidad Nacional de Catamarca

Maximio Victoria N° 55 - C.P: 4700 - San Fernando del Valle de Catamarca.

Tel: +54 (03834) 435112. Int 181

²Departamento de Informática/Facultad de Ciencias Físico, Matemáticas y Naturales/Universidad Nacional de San Luis.

Ejército de los Andes N° 950. San Luis. Tel: +54 (0266) 4520300. Int. 2127

ilazarte@tecno.unca.edu.ar, carlosacostap@tecno.unca.edu.ar,

gvilallo@tecno.unca.edu.ar/unsl.edu.ar

Resumen

En una colaboración inter-organizacional, las organizaciones se centran en el diseño de procesos (de negocio) colaborativos para acordar el comportamiento de la colaboración. No obstante, el diseño de procesos internos (de interfaz y de integración) constituyen un desafío importante para que las organizaciones puedan implementar y gestionar colaboraciones inter-organizacionales.

La línea de investigación presentada en este artículo se centra en la definición de métodos y herramientas que posibilitan el diseño y la gestión de los modelos de procesos de negocio internos con el propósito de integrar los mismos con los procesos de negocio colaborativos acordados en el marco de colaboraciones inter-organizacionales.

Por otro lado, en esta misma línea de investigación se busca definir metodologías, métodos y herramientas que permitan descubrir, monitorear y mejorar procesos de negocio reales de organizaciones tanto públicas como

privadas, mediante la aplicación de técnicas de minería de procesos sobre información de las actividades realizadas en cada proceso de negocio ejecutado en dichas organizaciones. Esta información, que es registrada por los sistemas de información, puede ser aprovechada para aprender sobre el comportamiento histórico de los procesos, y de esta manera detectar fuentes de problemas y oportunidades.

Palabras clave: Gestión de procesos de negocio, Colaboraciones inter-organizacionales, Minería de procesos, Mejora de procesos, Desarrollo dirigido por modelos

Contexto

Este trabajo está inserto en dos proyectos de investigación: el proyecto "Método para minería de procesos en entornos de colaboraciones inter-organizacionales" y el proyecto "Fortalecimiento de la calidad y la productividad en Ingeniería de Software".

El proyecto "Método para minería de procesos en entornos de colaboraciones inter-organizacionales" fue presentado y aprobado en la Convocatoria de Proyectos de Fortalecimiento de Redes Interuniversitarias IX, cuyo objetivo principal es promover la constitución y fortalecimiento de redes entre instituciones universitarias argentinas y extranjeras. En dicho proyecto participan docentes investigadores pertenecientes al Centro de Investigación y Desarrollo de Ingeniería en Sistemas de Información (CIDISI), dependiente de la Universidad Tecnológica Nacional, Facultad Regional Santa Fe (UTN-FRSF); de la Facultad de Tecnología y Ciencias Aplicadas (FTyCA) de la Universidad Nacional de Catamarca (UNCa); y de la Facultad de Ingeniería y Ciencias de la Universidad Autónoma de Tamaulipas, México.

El proyecto "Fortalecimiento de la calidad y la productividad en Ingeniería de Software", es evaluado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Catamarca y se desarrolla y ejecuta en ámbitos del Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas (FTyCA) de la Universidad Nacional de Catamarca (UNCa) y en el Departamento de Informática de la Facultad de Ciencias Físico, Matemáticas y Naturales de la Universidad Nacional de San Luis.

Introducción

La globalización, los mercados modernos, las nuevas filosofías de gestión de organizaciones y los avances en las Tecnologías de Información y Comunicación, alientan a las organizaciones a formar redes colaborativas y establecer colaboraciones inter-organizacionales entre los miembros

de dichas redes. Las colaboraciones inter-organizacionales están impulsadas por la necesidad de agilidad, adaptabilidad y flexibilidad de las organizaciones para mantener o mejorar su desempeño y competitividad en el mercado global [1,2,4].

Una colaboración inter-organizacional implica una integración orientada a procesos entre organizaciones heterogéneas y autónomas, que debe ser alcanzada tanto a *nivel organizacional* como a *nivel tecnológico*. A nivel organizacional, las organizaciones se centran en el diseño de procesos (de negocio) colaborativos para acordar el comportamiento de la colaboración. Un *proceso colaborativo* define una vista global de las interacciones entre las organizaciones para alcanzar metas comunes. Estos procesos colaborativos sirven como una base contractual para las colaboraciones inter-organizacionales, pero no son ejecutables. Para implementar y ejecutar un proceso colaborativo en forma descentralizada, cada organización debe definir sus procesos de negocio internos, denominados *procesos de interfaz* (públicos) y *procesos de integración* (privados). La solución definida en este nivel se denomina *solución inter-organizacional* [1].

A nivel tecnológico, las organizaciones generan las especificaciones (código) ejecutables de procesos de integración y las interfaces de los sistemas de las organizaciones usando estándares Business-to-Business. La solución definida en este nivel se denomina *solución tecnológica*. Esta solución permite dar soporte a la ejecución de los procesos colaborativos. En consecuencia, las soluciones definidas en ambos niveles deben tener una mutua correspondencia [1].

La solución inter-organizacional debe definirse usando modelos conceptuales para lograr el entendimiento y la comunicación de los procesos definidos entre todos los “stakeholders” participantes. La definición de estos modelos conceptuales de procesos con independencia de la tecnología de implementación posibilita su implementación en diferentes plataformas, facilitando y fomentando su re-uso [1].

Además, los modelos de procesos de interfaz e integración deben ser interoperables y mantenerse sincronizados y consistentes con los modelos de procesos colaborativos [1,4]. Por lo tanto, se requieren de métodos y herramientas que posibilitan el diseño y la gestión de los modelos que conforman la solución inter-organizacional.

Por otro lado, las tecnologías de información en general y los sistemas de información en particular desempeñan un rol importante en la gestión de procesos de negocio de organizaciones públicas o privadas, debido a que una gran cantidad de actividades que dichas organizaciones realizan son soportadas por los sistemas de información. Varios tipos de actividades contenidas en los procesos de negocio pueden ser ejecutadas automáticamente por los sistemas de información, sin la participación de un humano. La información sobre las actividades realizadas en cada proceso de negocio ejecutado son registradas por estos sistemas, la cual puede ser aprovechada para aprender sobre el comportamiento histórico de los procesos, y de esta manera detectar fuentes de problemas y oportunidades.

La minería de procesos es una disciplina emergente que busca analizar la información que los sistemas de información registran sobre los procesos de negocio, para poder entender,

monitorear, analizar y mejorar dichos procesos, mediante métodos y técnicas para proveen información basada en hechos para dar soporte a las mejoras de procesos. Esta nueva disciplina se construye sobre enfoques dirigidos por modelos de procesos y minería de datos [3].

Existen dos factores principales para el creciente interés en la minería de procesos. Por un lado, cada vez más eventos se registran, proporcionando información detallada sobre la historia de los procesos de negocio. Por otro lado, existe una necesidad de mejorar y dar soporte a los procesos de negocio en entornos competitivos y rápidamente cambiantes.

Por lo expresado, se requieren metodologías, métodos y herramientas que permitan descubrir, monitorear y mejorar procesos reales de organizaciones tanto públicas como privadas.

Líneas de Investigación, Desarrollo e Innovación

En la línea de investigación presentada en este trabajo se llevarán a cabo actividades relacionadas a las siguientes temáticas:

- Gestión de procesos de negocio
- Colaboraciones inter-organizacionales
- Minería de procesos
- Mejora de procesos
- Desarrollo dirigido por modelos

Resultados y Objetivos

El objetivo de esta línea de investigación es definir métodos y herramientas que posibiliten el diseño y la gestión de los modelos de procesos de negocio internos con el propósito de integrar los mismos con los procesos de

negocio colaborativos acordados en el marco de colaboraciones inter-organizacionales. Por otra parte, también se busca definir metodologías, métodos y herramientas que permitan descubrir, monitorear y mejorar procesos reales de organizaciones públicas y privadas a través de la extracción de conocimiento de los registros de los sistemas de información, aplicando técnicas de la minería de datos.

Formación de Recursos Humanos

El equipo de trabajo de esta línea de investigación cuenta con integrantes que se encuentran en la etapa de desarrollo de sus tesis de posgrado en el marco de la carrera de maestría en Ingeniería del Software en temas específicamente relacionados al área del proyecto. Además, los participantes se desempeñan como docentes de la carrera de Ingeniería en Informática (UNCa), e Ingeniería en Informática y Licenciatura en Ciencias de la Computación (UNSL). También se prevé la capacitación y formación de recursos humanos, que contempla las siguientes actividades:

- Participación en cursos de actualización y posgrado en el área de estudio.
- Participación en talleres o workshops de herramientas informáticas relacionadas con la temática de IS.
- Dirección y asesoramiento de tesinas de grado de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa.
- Dirección y asesoramiento de tesis de posgrado de la Maestría en Ingeniería de Software.
- Coordinar el dictado de cursos de capacitación según la demanda de

organismos gubernamentales competentes e industria del software.

- Transferencia de tecnologías de IS a la industria del software local.

Referencias

1. Lazarte, Ivanna M. *Diseño y gestión de modelos de procesos de negocio en colaboraciones inter-organizacionales*. Tesis Doctoral. Santa Fe. ISBN 978-987-33-3845-8.
2. Marlon Dumas, Marcello La Rosa, Jan Mendling, Hajo A. Reijers. *Fundamentals of Business Process Management*. Verlag Berlin Heidelberg, (2013). ISBN: 978-3-642-33142-8.
3. van der Aalst, Wil. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag Berlin Heidelberg, (2011). ISBN: 978-3-642-19344-6.
4. Weske, Mathias: *Business Process Management: Concepts, Languages, Architectures*. 2nd ed. Springer-Verlag Berlin Heidelberg (2012). ISBN: 978-3-642-28615-5.

Especificación de indicadores en el proceso de software con tecnologías libres

Mónica D. Tugnarelli, Cristian D. Pacifico, Martín M. Perez

Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos
Av. Tavella 1425, (E3202KAC) Concordia - ER, Argentina - Tel:(+54)(345)4231400
e-mail: {montug, cripac, marper}@fcad.uner.edu.ar

Resumen

Es indiscutible que el Software Libre ha cambiado la manera de entender las libertades de los usuarios y los procesos de desarrollo, permitiendo la creación de un modelo de negocio económico y rentable. Sin embargo, existen impedimentos a su adopción e implementación, siendo el más relevante la escasa información dirigida al nivel gerencial o profesional. Por tal motivo, actividades de observación tecnológica adquieren especial relevancia por ser el espacio para relevar datos y sistematizar indicadores que proporcionen información del estado, desarrollo y evolución de sectores involucrados con las TICs libres. En este trabajo se presentan los objetivos de un Observatorio Tecnológico a constituirse para tal fin; como así también se delinean algunos indicadores de medición de diferentes aspectos que permitan conocer el grado de impacto de las tecnologías libres en la región.

Palabras clave: software libre, observatorio, tecnologías libres, indicadores, vigilancia tecnológica

1. Contexto

La línea de I+D tiene como antecedente destacado la iniciativa de desarrollo tecnológico para el Fortalecimiento a la Innovación Tecnológica en Aglomerados Productivos **FONTAR FIT-AP 003/12** “Cluster de Software Libre” y, en particular, la **Acción 1:** “*CERTASOL - Centro de Referencia en Tecnologías Abiertas y Software Libre*”, acción llevada a cabo por la Facultad de Ciencias de la Administración de Universidad Nacional de Entre Ríos [21]. Esta unidad académica, varias empresas de desarrollo de software y la Cámara Argentina de Software Libre conforman este aglomerado

productivo exclusivamente dedicado a apoyar y fomentar el desarrollo e implementación de software libre. Entre los objetivos principales del FIT-AP se consideran la constitución de un Centro de Referencia para certificar software y tecnologías libres, así como profesionales y empresas que desarrollan y utilizan dicho software. En conjunción, también se establece un Observatorio Tecnológico para el relevamiento, análisis y seguimiento de información referida a la adopción de estas tecnologías.

El Proyecto **PID-UNER 7049** denominado “*Guías para la aplicación de normas de calidad para los procesos de ingeniería de software en productos desarrollados con lenguajes de programación open source: relevamiento y aplicación en PYMES de la zona de influencia de la UNER Concordia*”, se articula con la acción anteriormente descrita estableciéndose como ámbito científico-académico para el desarrollo y redacción de guías y normas para certificación de productos y procesos de Software Libre y Tecnologías Abiertas a utilizarse en el aglomerado y en otros sectores productivos. Además, para conocer el mercado potencial de aplicación de dicha asesoría tecnológica, es necesario relevar y analizar el grado de penetración y conocimiento del Software Libre y Tecnologías Abiertas en el sector productivo regional. Todo esto constituye información necesaria para poder planificar y dirigir el asesoramiento en la implementación de las guías y la medición del impacto de la adopción de las mismas. Estas dos últimas tareas son incumbencias del Observatorio y actividades del proyecto.

2. Introducción

Es indiscutible que el Software Libre es una revolución tecnológica que ha cambiado la manera

de entender las libertades de los usuarios y los procesos de desarrollo, permitiendo la creación de un modelo de negocio económico y rentable [6, 9, 5]. Si bien es una de las tecnologías más inclusivas y colaborativas que existen [19, 22], aún no ha logrado una amplia inserción en el mundo de los negocios en comparación, por ejemplo con su uso en el ámbito académico.

Una de las principales debilidades que afecta a las tecnologías libres, en cuanto a su adopción e implementación, es la escasa información dirigida al nivel gerencial o profesional, lo que facilita la creación de mitos que rodean al software libre en cuanto a su filosofía, al uso, a la calidad y al modelo de negocio que lo sustenta [2, 7, 10, 1]. Esta escasez de información, y más aun de *información sistematizada*, genera reticencia a la hora de tomar decisiones relacionadas a la elección o implementación de tecnologías dentro de una organización, optándose generalmente por adquirir herramientas tecnológicas propietarias a costos superiores, que ofrecen una mínima auditabilidad de los datos y con escasas posibilidades de personalización [17, 8].

Considerar un observatorio como fuente primordial para la captación y difusión de información, en este caso tecnológica, es una herramienta con varios antecedentes en el mundo. En Iberoamérica, existen destacados observatorios relacionados con el software libre [20]; ya sean iniciativas internacionales [12]; así como nacionales en España [3, 16, 15], Uruguay [11], México y Perú. En Argentina, la iniciativa más conocida es el Observatorio de la Cámara de Empresas de Software y Servicios Informáticos de la República Argentina (CESSI) que brinda servicios y reportes sobre la industria [4].

El Observatorio Tecnológico al que se hace referencia en el contexto de este trabajo propone constituirse como fuente de información confiable, actualizada y de libre acceso, con el objetivo primordial de crear un espacio adecuado de difusión y vinculación para tratar de lograr el aprovechamiento de las capacidades profesionales de la región y de las posibilidades que el uso de TICs libres trae aparejado, tanto en términos de desarrollo social, humano y económico, como de innovación para el sector de la industria del software. Todo esto se desarrolla en el área de influencia de nuestra Facultad -con sede en la ciudad de Concordia, Entre Ríos- teniendo en cuenta la potencial proyección a nivel nacional e internacional.

3. Línea de investigación, Desarrollo e Innovación

El proyecto PID-UNER 7049 descripto propone la creación de un marco de referencia para la aplicación de normas de calidad de productos desarrollados con lenguajes de programación *open source* y/o tecnologías libres, con el cual se intentará seleccionar normas específicas de calidad para procesos y productos de Ingeniería en Software, la especificación de propósitos, límites, y destinatarios de los modelos y guías de aplicación de las normas de calidad elegidas.

Este objetivo requiere de la conformación de una base de conocimiento de las organizaciones privadas y públicas que utilicen tecnologías libres, y de las empresas productoras de software que desarrollan su actividad en la zona de referencia de nuestra Facultad. En esta tarea, la misión del Observatorio adquiere especial relevancia por ser el espacio para relevar datos y sistematizar indicadores que proporcionen información del estado, desarrollo y evolución de sectores de la región involucrados con las TICs libres.

En esta línea de I+D para cumplir con la misión propuesta, se han definido los siguientes objetivos:

- Conocer el grado de utilización e innovación en TICs del sector productivo local y regional, mediante la obtención, reunión, sistematización y análisis de información específica
- Construir indicadores con el fin de obtener información cuantitativa y cualitativa, con especial relevancia aquellos que permitan referenciar las particularidades de la región.
- Relevar la oferta y demanda de productos y servicios del sector de software en su conjunto y, en particular, de TICs libres.
- Relevar la características de las empresas que desarrollan servicios TI basados en Software Libre y el modelo de negocio utilizado
- Relevar la demanda de recursos humanos calificados, y los perfiles requeridos por el mercado TIC regional.
- Relevar el estado de madurez de los procesos de desarrollo de productos de Software Libre.
- Publicar y difundir información, a fines de especializar el conocimiento de la situación ac-

tual como base de desarrollos futuros, identificación de oportunidades y nuevos campos de aplicación.

Según estos objetivos, y focalizando en el ámbito de Ingeniería en Software, es determinante la identificación de indicadores que midan la adopción de TICs libres en el proceso de construcción de software, en particular el uso de las herramientas de desarrollo basados en Software Libre; para esto, es necesario un estudio detallado de los procesos de desarrollo involucrados. Para establecer estos indicadores, se partirá del relevamiento de casos de éxito en el uso de esta tecnología buscando determinar los puntos comunes de las empresas relevadas y de los recursos humanos que intervienen en dichos procesos. De igual manera es importante especificar las diferencias entre los casos relevados.

Con esta finalidad, se establecerán facetas a investigar en los procesos de los casos elegidos. En particular, además del proceso de ciclo de vida de desarrollo, interesa conocer los procesos de: investigación e innovación, de capacitación al personal específico, y de su reclutamiento. Es importante conocer si existen planes de uso de Software Libre y planes de promoción y colaboración con proyectos libres como, por ejemplo: el reporte de errores, documentación colaborativa en proyectos libres, participación en comunidades de desarrollo, etc.

Teniendo en cuenta trabajos similares y de referencia [18, 3, 20, 14, 13], algunos de los indicadores que se han diseñados para conocer el grado de adopción de TICs libres son:

- Porcentaje de tareas de la organización que emplean herramientas informáticas.
- Porcentaje de software de base (servidores, gestores de contenido, bases de datos) que son de licenciamiento libre.
- Porcentaje de software de aplicaciones (suits de oficina, clientes de correo electrónico, herramientas de productividad) que son de licenciamiento libre.
- Porcentaje de software adquirido con licencia vs software gratuito (freeware) vs Software Libre.
- Estimado de costo anual por pago de licencias/actualizaciones.
- Cantidad de empleados específicos de los procesos de desarrollo de software.
- Grado de importancia de la capacitación de personal.
- Porcentaje de pericias/aptitudes relacionadas con TICs libres que tienen los empleados.
- Porcentaje de pericias/aptitudes relacionadas con TICs libres que se tienen en cuenta para el reclutamiento.

4. Resultados y Objetivos

Una de las actividades iniciales del proyecto es la realización de encuestas dirigidas a PYMES y emprendedores productores de software, como así también a organizaciones privadas y públicas para conformar una base de partida en la construcción de indicadores de medición de diferentes aspectos que permitan conocer el grado de impacto de las tecnologías libres en la región. De esta manera, se tendrá una fuente primaria para conocer, por ejemplo, el grado de implementación de TICs en la región, el modelo de negocios desarrollado por el sector informático, recursos humanos disponibles, perfil del profesional informático requerido, los emprendedores de la industria, entre otros.

Como resultado de las actividades proyectadas se espera alcanzar los siguientes objetivos:

- Elaborar, sintetizar y sistematizar indicadores de medición.
- Contar con estudios de mercado e informes sobre uso de TICs, ámbitos de aplicación y las capacidades técnicas requeridas en profesionales.
- Generar recomendaciones y servicios informativos sobre el uso de tecnologías específicamente para cada ámbito del sector productivo.
- Difundir experiencias, casos de éxito y buenas prácticas relacionadas con la aplicación de tecnologías libres.
- Incentivar la integración de organizaciones o comunidades de desarrollo con la universidad para fortalecer la colaboración e investigación lo que, a su vez, traería aparejada la formación de personal capacitado.
- Identificar y explorar tecnologías cuyo impacto es o será importante para los negocios, la industria y la sociedad en la región.

- Identificar los nuevos desafíos que deberán encararse en actividades de docencia, investigación, extensión e innovación en el área de las TICs.

A futuro, una gran oportunidad para profundizar la vinculación con el medio radica en que el Observatorio de Tecnologías Libres emprenda acciones hacia una vigilancia tecnológica del sector TICs libres, ofreciendo este servicio al público interesado y especialmente a las organizaciones y PYMES regionales. Si bien muchas organizaciones cuentan con sistemas de vigilancia tecnológica propios, es algo costoso de implementar, por ejemplo, para una PYME y además se requiere de personal dedicado y capacitado en la tarea. Estas condiciones, ocasionan que no todas las empresas y aún organizaciones de mayor tamaño puedan acceder a este sistema, cuestión que se ve agravada por el volumen y la velocidad con que se genera la información en un mundo globalizado e interconectado.

5. Formación de Recursos Humanos

Se pretende realizar la formación en investigación de 1 (un) becario de Iniciación en la Investigación afectado a este proyecto PID; 2 (dos) becarios con becas de formación de recursos humanos de la UNER, y se prevé el desarrollo de al menos 4 (cuatro) proyectos de Trabajo Final para la carrera Licenciatura en Sistemas. Además, se espera que 2 (dos) integrantes del equipo defiendan sus tesis de maestría durante el desarrollo del proyecto, en tanto que se estima que un tercer integrante, quien inició el cursado de la *Maestría en Sistemas de Información* dictado en la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos durante el año 2015, comience la elaboración de su tesis sobre el final del desarrollo del proyecto.

Referencias

[1] Berg, Ryan: *Comment: Myths and Misconceptions on Securing Open-source Software*. InfoSecurity, 2013. <http://www.infosecurity-magazine.com/view/33945/comment-myths-and-misconceptions-on-securing-opensource-software/>.

[2] Byfield, Bruce: *The Twelve Top Myths of Free/Open Source Software*. LinuxPlanet, 2008. <http://www.linuxplanet.com/linuxplanet/reports/6597/1>.

[3] CENATIC: *Observatorio Nacional de Software de Fuentes Abiertas. España*. <http://observatorio.cenatic.es>.

[4] CESSI, Cámara de Empresas de Software y Servicios informáticos: *Reporte semestral sobre el Sector de Software y Servicios Informáticos de la Argentina*. Reporte a diciembre 2012, Septiembre 2013. <http://www.cessi.org.ar/opssi/>.

[5] Cooper, Pete y Amir Nettle: *Free and open source software business and sustainability models*. OSS Watch, 2014. <http://oss-watch.ac.uk/resources/businessandsustainability>.

[6] Ferris, Michael: *Open source code and business models: More than just a license*. OpenSource.com, 2014. <http://opensource.com/business/13/5/open-source-your-code>.

[7] Ghosh, Rishab A.: *Understanding Free Software Developers: Findings from the FLOSS Study*. En Hissam, Scott y Karim Lakhami (editores): *Perspectives on Free and Open Source Software*. The MIT Press, Junio 2005.

[8] Heath, Nick: *Six open source security myths debunked - and eight real challenges to consider*. ZDNet, 2013.

[9] Ibañez, Luis: *Open source economic model: Sell the license or charge a consulting fee?* OpenSource.com, 2013. <http://opensource.com/education/13/2/open-source-economic-model>.

[10] Laguna, Rafael: *4 Myths About Open Source to Put to Rest*. Innovation Insights, 2013.

[11] ObservaTIC: *Observatorio de Tecnologías de Información y Comunicación*. Uruguay. <http://www.observatic.edu.uy>.

[12] ObservatorioCTS: *Observatorio Iberoamericano de la Ciencia, la Tecnología y la Sociedad de la OEI*. <http://observatoriocts.org>.

- [13] OECD: *Frascati Manual. Guidelines for Collecting and Reporting Data on Research and Experimental Development*. The Measurement of Scientific, Technological and Innovation Activities, 2015. [/content/book/9789264239012-en](#).
- [14] OECD y Eurostat: *Oslo Manual. Guidelines for Collecting and Interpreting Innovation Data, 3rd Edition*. The Measurement of Scientific, Technological and Innovation Activities, 2005. [/content/book/9789264013100-en](#).
- [15] OPTI, Fundación: *Observatorio de Prospectiva Tecnológica Industrial. España*. <http://www.opti.org>.
- [16] OVTT: *Observatorio Virtual de Transferencia de Tecnología. España*. <http://www.ovtt.org>.
- [17] Paget, Brian: *Viewpoint: 5 open source myths*. FederalTimes.com, 2013. <http://www.federaltimes.com/article/20131215/ADOP06/312150005/Viewpoint-5-open-source-myths>.
- [18] Presidencia de la Nación Argentina, Ministerio de Ciencia, Tecnología e Innovación Productiva de Argentina, Secretaria de Planeamiento y Políticas: *Sistema Integrado de indicadores CTI*. <http://indicadorescti.mincyt.gob.ar/>.
- [19] Rossi, Bruno, Barbara Russo y Giancarlo Succi: *Open Source Software and Open Data Standards as a form of Technology Adoption: a Case Study*. En Feller, Joseph y cols. (editores): *Open Source Development, Adoption and Innovation*, volumen 234 de *IFIP — The International Federation for Information Processing*, páginas 325–330. Springer US, 2007.
- [20] Tecnología (RICYT), Red Iberoamericana de Indicadores de Ciencia y: *Normalización de Indicadores de Innovación Tecnológica en América Latina y el Caribe*. Plan de Gobierno, Organización de Estados Americanos (OEA) - Programa CYTED COLCIENCIAS/OCYT, Marzo 2001. http://www.uis.unesco.org/Library/Documents/Bogota%20Manual_Spa.pdf.
- [21] Tugnarelli, M.D., C.D. Pacifico y M.M. Perez: *Modelos de certificación para Software Libre*. WICC 2014 XVI Workshop de Investigadores en Ciencias de la Computación, páginas 606–610, 2013. <http://hdl.handle.net/10915/41616>.
- [22] Wasserman, Tony: *Building a Business on Open Source Software*, Enero 2009.

Análisis sobre adopción de metodologías ágiles en los equipos de desarrollo en pymes del NEA

Mariano Rujana; Noelia Romero Franco; Nicolás Tortosa; Gabriela Tomaselli;
Noelia Pinto

GICS (Grupo de Investigación en Ingeniería y Calidad del Software), UTN,
FRRe

French 414, Resistencia, Chaco
{nicotortosa; gabriela.tomaselli; ns.pinto}@gmail.com

Resumen

Conocer qué tipo de metodologías de desarrollo de software están utilizando las empresas del NEA es fundamental para delinear y establecer estrategias que permitan la mejora de sus procesos y colaborar, así, en la obtención de productos software de calidad.

El presente artículo aborda de manera empírica el panorama respecto a la adopción de metodologías ágiles en equipos de desarrollo de software en empresas de Chaco y Corrientes. Se incluyen resultados logrados a partir del procesamiento de la información recabada y acciones futuras en base a las conclusiones obtenidas.

Palabras clave: Calidad del Producto y del Proceso de Software, Metodologías Ágiles, Pymes del NEA

Contexto

El trabajo que aquí se presenta está enmarcado en el proyecto “Framework para la evaluación de calidad de software”, que es financiado por la UTN y ejecutado en el Grupo de Investigación en Ingeniería y Calidad de Software (GICS) de la Facultad Regional Resistencia, con el código UTI2205TC.

Asimismo algunas actividades son compartidas con el Proyecto de Desarrollo Tecnológico y Social (PDTS) “Aporte a la

competitividad de las empresas de desarrollo de Software del NEA” [1], también radicado en el GICS y aprobado por el Consejo Interuniversitario Nacional (CIN) en su Convocatoria 2014.

Introducción

En Argentina, y según el último informe anual [2] del Observatorio Permanente de la Industria del Software y Servicios Informáticos de la República (OPSSI), este sector se conforma con más del 85% de empresas Pymes con menos de 200 trabajadores, donde se ubica el mayor índice de empleabilidad de la industria. Las mismas constituyen, hoy, un eslabón fundamental en el sector productivo de Argentina, por lo que resulta necesario implementar ciclos de mejora de procesos que contribuyan con el desarrollo de esta industria.

Es por ello que la calidad del producto final, los bajos costos y las entregas oportunas se transforman en elementos claves para el incremento de las ventas internas y la proyección a nivel internacional del sector. Para esto resulta fundamental la implementación de cuestiones asociadas a la gestión de calidad, no solamente desde el punto de vista del proceso de desarrollo sino también enfocada hacia el producto final.

En este sentido, varios autores [3][4][5] coinciden en que resulta difícil para las pymes

implementar programas de Mejoras de Proceso de Software (SPI) fundamentalmente por la falta de seguimiento de los planes de acción y de implantación debido al alto costo que significan. De esta forma, los parámetros de tiempos de desarrollo y costo de soluciones, afectarán directamente al trabajo que se realice, siendo la calidad la primera variable de ajuste disponible.

Así, existen numerosas propuestas metodológicas para el desarrollo de software que inciden en distintas dimensiones del proceso. Las propuestas más tradicionales se centran especialmente en una rigurosa definición de roles, de las actividades involucradas, los artefactos que se deben producir, y las herramientas y notaciones que se usarán [6]. Sin embargo, este enfoque no resulta ser el más adecuado para muchos de los proyectos actuales, donde el entorno del sistema es muy cambiante y en donde se exige reducir drásticamente los tiempos de desarrollo, pero manteniendo una alta calidad. Surgen, entonces, las metodologías ágiles, que persiguen principios como la entrega incremental de funcionalidad nueva al cliente, priorizándola según el valor de negocio que agrega (de esta forma el producto de software evoluciona en las diferentes entregas), mejora continua y el énfasis en la colaboración cercana entre el equipo de programadores y los expertos del negocio [7].

Las organizaciones como las pymes presentan características particulares que favorecen la implementación de las técnicas impulsadas por las metodologías ágiles, colaborando, también, en la búsqueda de la calidad orientada al proceso y producto de software [8].

Por otra parte, esas mismas características que definen las pymes como tales, son las que en muchos casos conspiran contra la adopción de políticas de Calidad, diseñadas en su mayoría para grandes empresas o equipos de desarrollo de software.

El análisis que se presenta aquí comprende los resultados obtenidos a partir de una

encuesta realizada en la Industria del Software del NEA respecto a la adopción de metodologías ágiles y el grado de implementación de cuestiones asociadas a la gestión de la Calidad.

Primero se presentan los ejes que se utilizaron para realizar el relevamiento inicial, y luego el análisis de la información recabada. Finalmente se exponen conclusiones y desarrollos futuros a partir de este estudio preliminar.

Líneas de Investigación, Desarrollo e Innovación

El objetivo general del Proyecto “Framework para la evaluación de calidad de software” es contribuir a la mejora en la calidad del software mediante modelos y métricas aplicados al producto y al proceso de desarrollo, como forma de incrementar la competitividad de quienes lo desarrollan y de aumentar la eficiencia, confiabilidad y seguridad en los distintos ámbitos de aplicación de sistemas informáticos. En el marco de este Proyecto algunas de las actividades que se realizan son:

- Diseño de encuestas a aplicar en empresas de los Polos IT Chaco y Corrientes para relevamiento de información respecto a implementación de Metodologías Ágiles y cuestiones asociadas a Calidad de Software.
- Aplicación y procesamiento de la encuesta.
- Análisis del procesamiento de resultados luego del relevamiento en las empresas.
- Estudio comparativo entre modelos de calidad que puedan implementarse y/o adaptarse en procesos ágiles.
- Propuesta de modelo de calidad aplicables a Pymes que utilizan metodologías ágiles o similar.

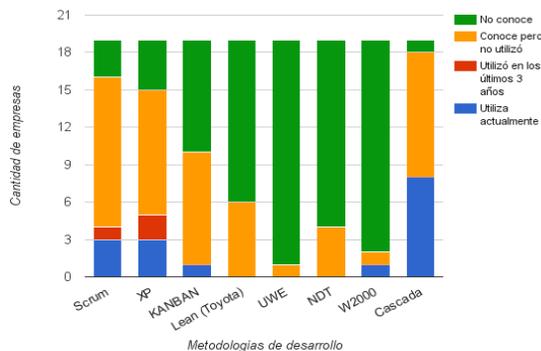
Así, para la actividad que contemplaba el diseño de la encuesta se tuvo en cuenta analizar diferentes ejes tales como:

Caracterización de la Empresa (tamaño, tipo, actividad principal, antigüedad en la Industria, etc), Metodologías de Desarrollo, Tecnologías, Gestión y Seguimiento de Proyectos y Calidad en el Proceso y Producto.

La encuesta fue aplicada en empresas del NEA, a través de un formulario on-line para facilitar los procesos de carga y posterior análisis de datos.

Resultados Obtenidos/Esperados

A partir del estudio y análisis de las encuestas realizadas a 19 empresas pymes de la región NEA, se observa que el 58% de las empresas encuestadas corresponden a empresas que cuentan con un área de Sistemas, mientras que el resto corresponde a empresas dedicadas al desarrollo de software. Se destaca



además que el 79% de las empresas tiene más de 5 años de antigüedad en ejercicio.

Figura 1. Relación Empresas-Metodologías de Desarrollo

Con respecto a las metodologías de desarrollo utilizadas por las empresas y si bien la gran mayoría admite conocer las metodologías de desarrollo ágiles más populares como Scrum o XP, tal como se muestra en la Figura 1, son pocas las que aplican actualmente o aplicaron en el último tiempo procesos ágiles; prefiriendo la mayoría volcarse a las metodologías tradicionales tal como el Desarrollo en Cascada.

Se observa asimismo poco y casi nulo conocimiento respecto a otras metodologías ágiles como KANBAN, Lean, etc.

Además un 26% de las empresas reconoce que utiliza una metodología propia, donde combinan principios ágiles con algunas cuestiones relacionadas a metodologías tradicionales.

Con respecto a la estimación de proyectos, 63% de empresas realizan alguna estimación en sus proyectos, teniendo en cuenta principalmente duración y costo de los mismos. De estas empresas, el 75% reconoció haber obtenido buenos resultados debido a la estimación.

Otro aspecto analizado fue la Gestión de Proyectos, observándose que la mayoría de las empresas encuestadas automatiza el seguimiento y administración de sus proyectos. En la Tabla 1 se incluyen los principales motivos por los cuales las empresas adoptan estas herramientas.

Motivo	Nro. de Empresas
Necesidad de realizar un seguimiento del desarrollo del proyecto	13
Necesidad de mejorar la coordinación del equipo de desarrollo	10
Incremento del tamaño y complejidad de los proyectos	5
Necesidad de optimizar el proceso de desarrollo	4
Necesidad de mejorar la coordinación del equipo de desarrollo	4
Ninguno o NC	4
Certificación ISO 9001	1

Tabla 1. Motivos por los cuales las empresas adoptan herramientas en la Gestión de Proyectos

Por otro lado, el 90% de las empresas no utilizan ninguna herramienta para la Gestión de Riesgos. En este sentido, en las Figuras 2, 3 y 4 se muestran las condiciones de finalización de los proyectos en las distintas empresas

teniendo en cuenta criterios de tiempo, presupuesto y funcionalidades originales:

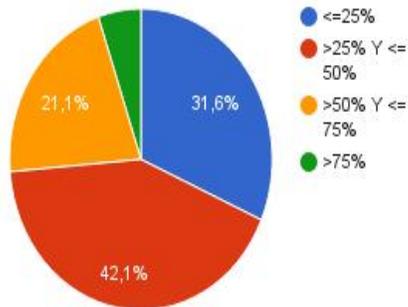


Figura 2. Porcentaje (%) de proyectos terminados a tiempo

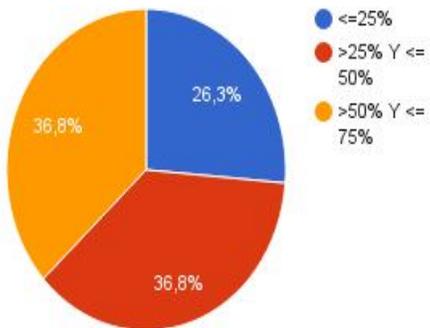


Figura 3. Porcentajes (%) de proyectos que finalizaron y están en funcionamiento

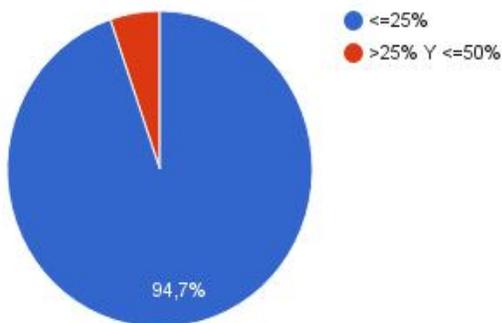


Figura 4. Porcentajes (%) de proyectos que se cancelaron antes de su terminación

Ninguna de las empresas entrevistadas realiza mediciones sistematizadas de sus productos de software, siendo el desconocimiento o el costo las principales barreras para realizarlas.

Respecto a Certificación de Calidad, sólo 4 cuentan con certificación ISO 9001. Resulta interesante destacar que dos de estas empresas certificadas implementan procesos ágiles para

el desarrollo y obtención de productos de software.

A partir del análisis realizado se concluye que las Pymes del NEA que han sido encuestadas comparten las siguientes cuestiones:

- La falta de conocimiento que impide implementación de metodologías ágiles en sus procesos de desarrollo.
- La mayoría ha adoptado herramientas de gestión de proyectos y utilización de estimaciones de los mismos.
- Implementación escasa o casi nula de Gestión de riesgos, Gestión de requisitos y medición de productos SW, siendo el desconocimiento o el costo las principales causas.
- La falta de certificaciones de calidad o interés de obtener certificaciones en el futuro.

En función a estas conclusiones, se ha iniciado el estudio comparativo de modelos de calidad que puedan aplicarse a procesos ágiles de desarrollo de Software en empresas PyME, y analizar la posible adecuación de estos modelos a las particularidades de este tipo de organizaciones.

Como acciones futuras se prevé trabajar en el diseño de un Framework que permita evaluar la calidad de procesos de desarrollo de Software basados en metodologías ágiles, y adecuarlo a la realidad de empresas pymes del NEA, a través del trabajo conjunto entre Universidad e Industria. El principal desafío será lograr involucrar y hacer partícipe del proyecto a todo el staff de las empresas que acepten trabajar en esta línea, ya que serán los ejecutores de las prácticas elegidas para ser implementadas.

Formación de Recursos Humanos

El equipo de trabajo del Grupo de Investigación en Ingeniería y Calidad de Software (GICS) de UTN Facultad Regional Resistencia está integrado por la Directora (Especialista, Categorizada en el Programa de

Incentivos y como Docente Investigadora de UTN), un Co-Director (Doctor, Categorizado en el Programa de Incentivos y como Docente Investigador de UTN) y dos Docentes Investigadores (Ingenieros, una Categoría “E” UTN y otro Categoría “G” UTN). Actualmente existen dos Tesis de posgrado radicadas en el Grupo.

Además el equipo no sólo incluye investigadores formados sino que incorpora y forma jóvenes investigadores algunos de los cuales colaboran como docentes en diversas cátedras y otros se encuentran insertos laboralmente en Pymes de Software del NEA. Se cuenta, así, con un Becario BINID, un Becario alumno de Investigación y dos alumnos de Ingeniería en Sistemas de Información que realizan su Práctica Supervisada en este Grupo.

Referencias

[1] Acuña, C., Cuenca Pletsch, L., Pinto, N., Tomaselli, G., Tortosa, N. (2015) “La vinculación Universidad-Industria como instrumento de mejora para la competitividad de las empresas de desarrollo de Software del NEA” 9º JUI. ISSN: 2451-7505

[2] “Reporte anual sobre el Sector de Software y Servicios Informáticos de la República Argentina” (2015). Disponible en <http://www.cessi.org.ar/opssi>.

[3] Pino, F., F. Garcia, M. Piattini (2006). “Revisión sistemática de mejora de procesos software en micro, pequeñas y medianas empresas.” Revista Española de Innovación, Calidad e Ingeniería del Software. (REICIS). Vol. 2(1) Abril pp. 6-23.

[4] Mas A., Amengual E. (2005). “Las mejoras de los procesos de Software en las pequeñas y medianas empresas (pymes). Un nuevo modelo y su aplicación a un caso real”.

Revista Española de Innovación, Calidad e Ingeniería del Software, Vol.1, No. 2

[5] Pasini, A. C., Esponda, S., Bertone, R. A., & Pesado, P. (2008). “Aseguramiento de Calidad en PYMES que desarrollan software.” XIV Congreso Argentino de Ciencias de la Computación.

[6] Letelier, P., Penadés, P. (2006) “Metodologías ágiles para el desarrollo de software: eXtreme Programming (XP)” Técnica Administrativa, Buenos Aires. ISSN 1666-1680

[7] Alliance, A. (2001). “Agile manifiesto”. Información disponible en <http://www.agilemanifesto.org>

[8] Muñoz, M., Casca, G., Valtierra, C. (2014) “Caracterizando las necesidades de las pymes para implementar Mejoras de Procesos Software: Una comparativa entre la Teoría y la Realidad”. Revista Ibérica de Sistemas y Tecnologías de la Información. ISSN 1646-9895.

Métodos y herramientas de estimación, gestión cuantitativa de proyectos, trazabilidad de requerimientos y entrega continua, orientados a la mejora de la calidad del software

Dapozo, Gladys N.; Greiner, Cristina; Irrazabal, Emanuel; Medina, Yanina; Ferraro, María; Lencina, Andrea; Chiapello, Jorge; Mascheroni, Agustín

Departamento de Informática. Facultad de Ciencias Exactas y Naturales
y Agrimensura. Universidad Nacional del Nordeste
{gndapozo; cgreiner, ferraro, yanina}@exa.unne.edu.ar
gaspagu.3@gmail.com; emanuelirrazabal@gmail.com

Resumen

Este proyecto de investigación busca contribuir a la mejora en la calidad del desarrollo y del producto software. La automatización de los métodos de estimación permitirá a las empresas mejorar la precisión de la duración y costo de los proyectos de desarrollo, generando además información histórica que retroalimenta a los métodos de estimación. La gestión cuantitativa, en particular, la gestión de incidentes y bugs, contribuirá a la eficiencia en los procesos de desarrollo. La sistematización de mecanismos de trazabilidad de los requerimientos y la incorporación de procedimientos que agilicen la puesta en producción de productos software seguros, contribuirán a mejorar la mantenibilidad del software, siendo este atributo uno de los más crítico en los ámbitos reales de producción de software.

Palabras clave: Estimación software. Incidencias. Entrega Continua. Gestión de Requerimientos.

Contexto

Las líneas de I/D corresponden al proyecto PI-F10-2013 “Métodos y herramientas para la calidad del software”, acreditado por la Secretaría de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE) para el periodo 2014-2017.

Introducción

Estimación

El desarrollo de software requiere de la estimación para controlar y administrar los recursos antes y durante el proyecto. Existen diferentes modelos de estimación que van desde las técnicas orientadas a los procesos hasta métodos paramétricos o algorítmicos, basados en datos históricos utilizados para definir y calibrar el modelo [1] [2]. La técnica (o conjunto de técnicas) que se utiliza debe adaptarse a los datos disponibles y la naturaleza del problema en la estimación. En [3] se concluye que a mayor especificidad del método en cuanto al contexto de desarrollo, mejor es la precisión de la estimación de esfuerzo y duración.

Asimismo, la determinación de costos de un producto es un aspecto clave para su comercialización. Un cálculo adecuado permite ganar clientes asegurando la expansión de las empresas. Considerando las características del proceso, del equipo, del producto y de la empresa, el Centro de Estudios de Ingeniería de Software de la Universidad de la Frontera (CEIS-UFRO) desarrolló un método que partiendo de una especificación de requerimientos basada en casos de uso y la productividad, llega a definir el costo de un producto de software transaccional web [4].

La realidad de las empresas de la región muestra, según un estudio cuyos resultados se detallan en [5], que las áreas

o empresas de sistemas, en gran medida, desconocen los métodos y herramientas orientados a mejorar la calidad del desarrollo y del producto software, siendo necesario realizar desde las universidades o los polos tecnológicos actividades de divulgación o capacitación sobre estos temas.

Entrega continua

La Entrega Continua de Software, en inglés Continuous Delivery (CD), se define como un enfoque en el cual los equipos mantienen la producción de software en ciclos cortos de tiempo, asegurando que el producto pueda ser lanzado de manera fiable en cualquier momento [6]. La idea es poder lanzar a producción un producto software libre de defectos con solo apretar un botón [7].

Existe un concepto similar al CD, que es el de Despliegue Continuo, en inglés Continuous Deployment (DC). El DC es una actividad que consiste en lanzar cambios continuamente al ambiente de producción [8]. La principal diferencia se encuentra en la fiabilidad a la hora de lanzar una nueva versión del producto: el DC busca integrar código a producción una, dos, y muchas más veces en el mismo día, en cambio, el CD se centra en hacerlo con la certeza de que el producto que se está lanzando a producción tiene un alto grado de calidad y se encuentre libre de defectos.

En el intento de implementar el CD, muchas organizaciones terminaron alcanzando solamente el DC. Por ejemplo, algunas empresas como Facebook, Atlassian, IBM, Adobe, Microsoft y Flickr, han tenido éxito en la implementación de diferentes enfoques para realizar entregas en periodos cortos de tiempo. Sin embargo, diferentes autores señalan que el proceso de lanzar rápidamente productos software a producción, aún presentan inconvenientes por resolver [9].

Uno de los principales problemas estaría en la calidad del producto software. Ésta puede disminuir, dado que,

al realizarse los despliegues del sistema con mayor frecuencia, aparecen más defectos en el producto [7]. Por tanto, es esencial desarrollar un enfoque de priorización de los diferentes aspectos en la calidad del producto software, teniendo en cuenta la forma de trabajo actual de las empresas de desarrollo software.

Para adoptar este enfoque, en la literatura se utiliza lo que se conoce como Tubería de Despliegue (DP - Deployment Pipeline). El DP es un estándar para automatizar el proceso de CD [7]. A pesar de que cada organización puede variar en la implementación de este estándar, el mismo se conforma de las siguientes actividades: Instalación, Compilación, Pruebas de Aceptación, Pruebas de Capacidad, Pruebas Manuales, Liberación a Producción.

Con este tipo de solución se logra un entorno automatizado de compilación y pruebas de código dividido en etapas, que generan resultados rápida y eficientemente. Sin embargo, aún no se han encontrado evidencias de un proceso formal y estandarizado, ni se han evidenciado diferentes infraestructuras teniendo en cuenta el tipo de sistemas o la prioridad en el tipo de pruebas [10].

Gestión cuantitativa - Incidencias

La gestión cuantitativa en el desarrollo de software proporciona una visión del grado de cumplimiento de metas, así como las causas que explican desviaciones significativas en procesos o productos [11]. La gestión de proyectos en base al conocimiento cuantitativo contribuye a la determinación de los aspectos de mayor relevancia, cuyo rendimiento afecta en forma significativa al logro de los objetivos del proyecto y la satisfacción de los clientes, obteniendo productos con un mayor nivel de calidad. La medición de código fuente aporta a este conocimiento, por ejemplo estableciendo relaciones en cuanto a la complejidad del software y la propensión a fallos, lo cual ofrece indicios sobre qué

clases o porciones de código debe enfocarse la prueba de software.

Vinculado con el código fuente aparece el concepto de deuda técnica, la que se contrae cuando las actividades relacionadas con el desarrollo de software no se realizan con los niveles de calidad adecuados y por tanto puede disminuir la mantenibilidad del código u ocasionar mayores costos en el proyecto [12].

A medida que el proyecto crece en cantidad de requerimientos, tamaño del código o personas involucradas, el éxito resulta cada vez más difícil de alcanzar. Es importante que los equipos se mantengan dentro de presupuestos y tiempos planificados. Una forma efectiva para alcanzar las metas es que los equipos cuenten con herramientas que les permitan sistematizar aspectos tales como peticiones de cambios, incidencias y bugs. SWEBOK define como petición de cambio (ChangeRequest, CR) a: “una solicitud para ampliar o reducir el alcance del proyecto; modificar las políticas, procesos, planes o procedimientos; modificar los costos o presupuestos; o revisar los horarios. Una fuente de CR es el inicio de medidas correctivas en respuesta a informes de problemas [13].

La International Software Testing Qualifications Board (ISTQB) señala que durante la ejecución de una prueba de software puede ocurrir que los resultados reales varíen de los esperados. En este caso se dice que existen incidentes, bugs, defectos o problemas. Sin embargo, marca diferencia entre incidentes y defectos o errores. Un incidente es cualquier situación en la que el sistema muestra un comportamiento cuestionable, y se denomina bug sólo cuando es la causa de algún problema que se está probando [14].

SWEBOK define un bug como “un defecto en el código fuente. Un paso, proceso o definición de datos incorrectos en el programa. La codificación de un error humano en código fuente” [13].

En la literatura especializada surge que la asignación de bugs es un aspecto crítico. En [15] se describe cómo la tarea de asignar un CR puede ser costosa en tiempo, y la reasignación doblemente costosa, señalando la falta de uso de una herramienta automatizada que asista en esta tarea, y la necesidad de considerar la información histórica para tomar decisiones efectivas, tales como la carga de trabajo actual y conocimiento de habilidades de sus desarrolladores. En [16] se describe la asignación de bug como un proceso social debido a la complejidad que implica determinar la persona más adecuada para resolver el problema, señalando la importancia de tener una base de conocimiento sobre la experticia de los desarrolladores y herramientas que abarquen información socio-técnica de la organización. En [17] se menciona el conocimiento de los desarrolladores y la repercusión en la carga de trabajo del equipo como desafíos claves en la asignación de un bug.

Trazabilidad

Los procesos de evolución del software varían considerablemente dependiendo del tipo de software a mantener, los procesos de desarrollo utilizados en una organización y el personal implicado en el proceso. Por tanto, es imprescindible contar con una herramienta de gestión de cambios que acompañe este proceso, para minimizar el impacto de los cambios en el sistema y lograr que los mismos se realicen de manera planificada y controlada. Se reconoce que realizar el seguimiento a los requisitos a lo largo del proceso de desarrollo de software no es tarea fácil. Todo artefacto de software cambia en el tiempo por la evolución en las necesidades de los usuarios. Para minimizar el impacto causado por dicha evolución, la práctica de la trazabilidad ha sido estudiada e implementada con diferentes modelos y técnicas que permiten lograr mayor calidad en los productos de software [18]. El desafío es sin duda, acompañar la gestión de

cambios con el suficiente nivel de trazabilidad que logre el equilibrio entre calidad de software y tiempo dedicado a la gestión.

Líneas de investigación y desarrollo

En la línea de Estimación se propone:

- En base a los resultados obtenidos en [3] se desarrollará una herramienta para automatizar la estimación de costos en proyectos web aplicando el método de estimación CEIS-UFRO.

En la línea de Entrega Continua:

- Estudiar, a partir de encuestas, el grado de importancia que los líderes de proyectos le confieren a requerimientos no funcionales.
- Realizar una revisión sistemática sobre modelos de entrega continua, haciendo hincapié en las estrategias, en el formato y tipo de pruebas y su automatización, y en los tipos de sistemas.
- Analizar características de las herramientas de entrega continua existentes, y contrastarlas con las necesidades de los profesionales, teniendo en cuenta el grado de importancia dado a los requerimientos no funcionales y los tipos de pruebas que se incorporan en los proyectos.

En la línea de Gestión Cuantitativa e Incidencias se propone:

- Un estudio detallado de herramientas de gestión de incidentes que se utilizan en empresas de software de la región NEA, identificando sus dificultades y aportes cuantificables en el desarrollo de software.
- Diseñar y construir una herramienta que permita realizar el seguimiento de las incidencias de los proyectos de software, aportando información de las características socio técnicas de los equipos involucrados. La herramienta permitirá gestionar un histórico de las incidencias, desde su reporte inicial hasta su corrección, siendo configurable

el ciclo de vida de las mismas para que se adapte a las necesidades de la organización. A su vez, tendrá en cuenta la experiencia, carga de trabajo actual e involucramiento de los desarrolladores.

En la línea de Trazabilidad, se propone:

- Un estudio detallado de herramientas de gestión de requerimientos que se utilizan en las empresas de software de la región NEA, desde el aporte de trazabilidad visible y cuantificable en el desarrollo de software.
- Proponer técnicas de trazabilidad como un conjunto de buenas prácticas en la gestión de requerimientos que contribuyan a la calidad del software sin sobrecargar las tareas de gestión.

Resultados y Objetivos

Los principales resultados de las actividades desarrolladas en estas líneas son:

En la línea de Estimación:

- Se evaluaron diferentes métodos de estimación, Puntos de Casos de Uso, CWADEE, Webmo y RESC, para estimar duración y esfuerzo, y se concluye que a mayor especificidad del método (en cuanto al tipo de desarrollo), mejor es el ajuste a la duración real. Como así también, a mayor disponibilidad y pertinencia de datos históricos de proyectos, mejor es la precisión de la estimación [3].
- Basado en los resultados obtenidos en [3], se desarrolló una herramienta que automatiza la estimación en los proyectos web aplicando los métodos estudiados. Este desarrollo fue parte del Proyecto Final de un alumno de la carrera.
- Analizados los datos de una encuesta realizada a empresas de software y áreas de sistemas de las organizaciones, respecto de la estimación se concluye que en la mayoría de las áreas/empresas las prácticas de estimación más utilizadas son la de juicio de experto y por analogía. Se conocen, pero no se

utilizan, o se utilizan poco, las técnicas paramétricas que pueden aportar mayor precisión en la estimación. Así también es bajo el porcentaje de las áreas/empresas que utilizan datos históricos para la estimación [19].

En la línea de Entrega Continua:

- Se realizó una Revisión Sistemática de la Literatura sobre la refactorización de software basada en valor para obtener conclusiones acerca de la manera en la cual se priorizan las mejoras del código fuente y del tipo de herramientas utilizadas. Los resultados se publicaron en [20].

- Asimismo, se realizó un estudio de la mejora de la mantenibilidad. En el artículo se presentaron los resultados de la utilización del modelo de medición y de las herramientas durante diez meses en una empresa española. Se detalla la evolución de las mediciones, la satisfacción de los desarrolladores y cómo han ido evolucionando las incidencias encontradas en las pruebas de aceptación [21].

- Se realizó una investigación preliminar para determinar el grado de importancia conferido a los atributos de calidad, mediante una encuesta a los responsables de aseguramiento de la calidad de los proyectos en centros de la Universidad de Ciencias Informáticas en Cuba. Se concluyó que la funcionalidad es la característica percibida como más importante. Esto tiene relación directa con los tipos de pruebas que se realizan de manera habitual antes de una entrega de software: las pruebas funcionales [22].

En gestión cuantitativa:

- Se realizó una Revisión Sistemática de la Literatura (RSL) sobre herramientas para el apoyo a la gestión de proyectos, la que permitió observar que dentro de los aspectos tratados con mayor frecuencia aparece muy fuerte la administración de los equipos de desarrollo, a través de herramientas de coordinación de actividades [23].

- Se profundizó el estudio de deuda técnica, con el propósito de proponer buenas prácticas para prevenir, medir y gestionar la deuda técnica en el desarrollo de software [24].

- Se abordó el tema de la priorización de las mejoras aplicadas al código fuente y el tipo de herramientas utilizadas, a través de una RSL [21].

En la línea de trazabilidad:

- En la encuesta realizada para determinar el estado de situación del desarrollo de software en la ciudad de Corrientes, considerando las empresas de software y las áreas de sistemas de las organizaciones, se determinó que es escasa o casi nula la presencia de trazabilidad de requerimientos, sobre las distintas herramientas de gestión utilizadas [19].

Formación de recursos humanos

En el Grupo de Investigación sobre Calidad de Software (GICS) están involucrados 5 docentes investigadores, 2 becarios de investigación de pregrado y 1 tesista de doctorado. Un alumno de la carrera finalizó la misma con un proyecto vinculado con estos temas.

Referencias

- [1] J. Andrés, D. Fernandez-Lanvin, P. Lorca. "Cost estimation in software engineering projects with web components development".2015, vol.82, n.192
- [2] E. Mendes, "Using knowledge elicitation to improve Web effort estimation: Lessons from six industrial case studies," Software Engineering (ICSE), 2012.
- [3] G. Dapozo, Y. Medina, A. Lencina, G. Pedrozo Petrazzini. "Métodos de estimación de esfuerzo y duración en proyectos web pequeños". Revista Eletrônica Argentina-Brasil de Tecnologias da Informação e da Comunicação. v. 1, n. 2 (2014)
- [4] J. E. Díaz Villegas, G. Robiolo. "Método de estimación de costos de un producto de software Web". 43 JAIIO - ASSE 2014 - ISSN: 1850-2792
- [5] G. N. Dapozo; C. L. Greiner, E. Irrazábal, Y. Medina, M. A. Ferraro, A. B.

- Lencina "Características del desarrollo de software en la ciudad de Corrientes".CACIC 2015.
- [6] L. Chen, "Continuous Delivery: Huge Benefits, but Challenges Too" in IEEE Software 03/2015. V. 32(2).
- [7] J. Humble and D. Farley. "Continuous delivery: reliable software releases through build, test, and deployment automation", 1st ed. Boston, US: Pearson Education, 2010.
- [8] H. H. Olsson, H. Alahyari, and J. Bosch. "Climbing the 'Stairway to Heaven'-A Multiple-Case Study Exploring Barriers in the Transition from Agile Development towards Continuous Deployment of Software" in 38th EUROMICRO Conference, 2012, pp. 392-399.
- [9] B. Fitzgerald and K. Stol, "Continuous Software Engineering and Beyond: Trends and Challenges," in 1st International Workshop on Rapid Continuous Software Engineering, 2014.
- [10] O. Prusak."Continuous Testing: The Missing Link in the Continuous Delivery Process". Blaze Meter. 2015.
<https://blazemeter.com/blog/continuous-testing-missing-link-continuous-delivery-process>
- [11] Gou, L., Wang, Q., Yuan, J., Yang, Y., Li, M., & Jiang, N. "Quantitative defects management in iterative development with BiDefect". Software Process Improvement and Practice, 14(4), 227-241. 2009.
- [12] W. Cunningham, "The wycash portfolio management system," in OOPSLA '92: Addendum to the proceedings on Object-oriented programming systems, languages, and applications (Addendum), 1992.
- [13] P. Bourque y R. E. Fairley, SWEBOK, Tercera ed., IEEE, 2014, pp. 82-87.
- [14] International Software Testing Qualifications Board (ISTQB). "What is an Incident in software testing?". [En línea].<http://istqbexamcertification.com/what-is-an-incident-in-software-testing/>
- [15] P. Anselmo da Mota y Y. Cerqueira. "Towards Understanding Software Change Request". Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering, 2013.
- [16] P. Guo y T. Zimmermann. "Not my bug! and other reasons for software bug report reassignments". CSCW '11 Proceedings of the ACM 2011 Conference on, 2011.
- [17] G. Bortis y A. van der Hoek. Teambugs: a collaborative bug. CHASE '11 Proceedings of the 4th International, pp. 69-71, 2011.
- [18] M. Tabares, f. Arango, R. Anaya. Una Revisión de Modelos y Semánticas Para la Trazabilidad de Requisitos. Revista EIA, ISSN 1794-1237 Número 6, p. 33-42. Diciembre 2006. Escuela de Ingeniería de Antioquia, Medellín. Colombia.
- [19] G.N. Dapozo, Y. Medina, A. Lencina. "La práctica de la estimación en empresas y áreas de Sistemas". Revista Eletrônica Argentina-Brasil de Tecnologias da Informação e da Comunicação. v.1, n. 4.2015
- [20] E. Irrazábal, C. Greiner, G. Dapozo. "La refactorización de software basada en valor: Revisión sistemática de la literatura". JAIIO 2015, ASSE, pp. 130 -144.
- [21] E. Irrazábal, "Mejora de la mantenibilidad con un modelo de medición de la calidad: resultados en una gran empresa". XXI CACIC. 2015.
- [22] M. Pérez García, E. Irrazábal, R. Carrasco-Velar, Y. Coca Bergolla. "Importancia de los requisitos no funcionales: estudio preliminar en una Universidad de Cuba". VII taller internacional de calidad en las tecnologías de la información y las comunicaciones calidad 2016. La Habana, Cuba (pendiente de publicación).
- [23] G. Dapozo, C. Greiner, J. Chiapello.. "Revisión Sistemática de Herramientas de Apoyo a la Gestión Cuantitativa de Proyectos de Software". III Jornadas de Investigación en Ingeniería del NEA y Países Limitrofes. 2014.
- [24] G. Dapozo, C. Greiner, E. Irrazabal, J. Chiapello. "Análisis de código fuente orientado a la calidad del producto". XVII Workshop de Investigadores en Ciencias de la Computación (WICC2015).

Framework de Evaluación de Productos *Software*

Paula Angeleri, Rolando Titiosky, Jorge Ceballos

Facultad de Ingeniería y Tecnología Informática/ Universidad de Belgrano

Av. F. Lacroze 1947, CABA

Paula.angeleri@ub.edu.ar, {paula.angeleri; rolando.titiosky; jorge.cebillos}@comunidad.ub.edu.ar

Resumen

El objetivo de este artículo es presentar la situación actual y los avances realizados en el proyecto de investigación MyFEPS *Metodologías y Framework para la Evaluación de Productos de Software* [1], desarrollado en la Facultad de Ingeniería y Tecnología Informática de la Universidad de Belgrano. En este contexto se describen las actividades de transferencia y ajustes al Framework, producto de la última fase del proyecto.

Palabras clave: Calidad de software, Evaluación de calidad de producto *software*, Framework MyFEPS.

Contexto

Evaluación de Productos *Software*

A partir del interés del IRAM¹, UNIT², ABNT³, INN⁴ y del BID⁵ de elaborar normas técnicas que permitiesen mejorar la competitividad de las empresas, nace el proyecto BID-FOMIN que incluye un Programa “Software” gracias al cual se adoptaron normas

ISO/IEC⁶ en Argentina y a Nivel MERCOSUR. A partir de la publicación de estas normas, IRAM desarrolla el servicio de certificación de productos software [2]. Para dar apoyo a este servicio, para especificar un modelo de calidad actualizado [3], y un proceso de evaluación que permitiese la ponderación de características de calidad en base a objetivos de evaluación [4], entre otros objetivos explicados anteriormente [5], [6], la Universidad de Belgrano inicia un proyecto de investigación para elaborar un Framework que facilite la evaluación de productos *software*: el proyecto MyFEPS. El proyecto fue financiado por la Universidad de Belgrano, y con aportes de la empresa TSOFT, a través de un aporte del FONTAR [7].

Introducción

El proyecto MyFEPS se llevó a cabo conforme a lo planificado, y cumplimiento sus objetivos [5], [6]. En primer lugar se elaboraron las bases teóricas [1], luego el modelo de calidad de productos software

¹ Instituto Argentino de Normalización y Certificación.

² Instituto Uruguayo de Normas Técnicas.

³ Asociación Brasileira de Normas Técnicas.

⁴ Instituto Nacional de Normalización de Chile.

⁵ Banco Interamericano de Desarrollo.

⁶ Las normas ISO de Tecnología de la Información son normas ISO/IEC porque dependen del JTC1, Comité Técnico Conjunto de la *International Standardization Organization* y de la *International Electro Technical Commission*.

QSAT⁷ [3], posteriormente el proceso de evaluación [4], basado en las normas internacionales ISO/IEC 14598 [12] e ISO/IEC 25040 [13], y mientras tanto se construían herramientas de apoyo [7], entre otros objetivos específicos. A continuación se llevaron a cabo tres proyectos para validar la adecuación del proceso de evaluación y del modelo de calidad. Estos fueron los proyectos *ConexionUB*, *eCOMM*, y *SHG* [6] [7], los cuales cumplieron sus objetivos, validando que el proceso es fácil de utilizar y adecuado, y que sus atributos y métricas referidas a las características básicas “usabilidad objetiva”, “satisfacción subjetiva de los usuarios”, y “facilidad de instalación”, permitieron medir su calidad (con algunos ajustes que surgieron de estos proyectos).

A la fecha, el Laboratorio de Testing de la Universidad de Belgrano ha realizado otros proyectos de evaluación de calidad utilizando MyFEPS, durante 2014 y 2015, y brindado cursos de capacitación, los que se describen a continuación:

Evaluación de Morpho

Una de las experiencias más importantes por las dimensiones de la evaluación y la cantidad de recursos involucrados, fue el proyecto Morpho. Proyecto que culminó con la certificación IRAM Chile del software desarrollado por la empresa Morpho para el Gobierno de Chile, Ministerio del Interior. Dadas las dimensiones del proyecto, IRAM Central asignó el trabajo a tres laboratorios, manteniendo la gestión de proyecto. Los laboratorios participantes fueron Laboratorio de Testing San Luis (LTSL), Laboratorio de Testing de Universidad de

Belgrano, y Laboratorio de Testing de la Universidad de Morón. A su vez LTSL coordinaba el trabajo de los tres laboratorios. El Proceso de Evaluación de la Calidad utilizado fue el mismo para todos los laboratorios involucrados, se usó MyFEPS para definir las etapas y las estimaciones de costo. El Laboratorio de Testing de UB usó también el proceso de evaluación MyFEPS para estimar el calendario y la planificación de recursos. Los otros laboratorios usaron métodos heurísticos propios para la definición de roles, puesto que la ISO no define un procedimiento de evaluación, sino el proceso y sus requisitos. El Modelo de Calidad elegido fue el de la norma ISO/IEC 25010 [14]. Las Características Evaluadas para los Módulos de Bloqueo, Inventario, Entrega, y Correspondencia, fueron: Adecuación funcional, Seguridad, Usabilidad. Al Modulo Correspondencia, se le agregó la característica Compatibilidad. Las Métricas utilizadas fueron en casi todos los casos ISO/IEC, excepto en algunas subcaracterísticas donde se recomendó el uso de las métricas de QSAT, dado que las métricas de la serie ISO/IEC 25000 [15] aún no se han publicado y se estaba trabajando con las métricas correspondientes a la ISO/IEC 9126 [16], que no las incluía.

Evaluación de Learning System

Uno de los proyectos fue Learning System, con el objetivo de evaluar la mejor plataforma de E-learning y su integración con un CRM orientado al negocio Educativo. Experiencia que permitió validar los atributos y métricas de varias características básicas. En el proyecto participaron un Project Manager, un Ingeniero de Software y tres Testers, y se evaluaron:

- LMS/LCMS: Ilias, Chamilgo y Moodle.

⁷ QSAT: sigla proveniente del apellido de sus creadores: Sorgen, Angeleri y Titiosky. La Q es por “quality”.

- CRM: SugarCRM, vTiger, Salesforce

Se Planificó, Cuantifico, Ejecuté y Concluyó la Evaluación desde la óptica de diferentes Stakeholders:

- LMS: Alumnos, Docentes, Responsable de Área.
- CRM: Alumnos, Administrador, Cliente, Sistema.

El Proceso de Evaluación de la Calidad utilizado fue MyFEPs con el que se definió las etapas, las estimaciones de costo, los tiempo necesarios y los recursos humanos y sus roles. El Modelo de Calidad elegido fue el QSAT. Las Características Evaluadas para ambos sistemas fueron: Adecuación Funcional, Eficiencia, Facilidad de Instalación, Seguridad de datos y de procesos, y Calidad de Artefactos. Las Métricas y sus Procedimientos fueron en todos los casos QSAT. La totalidad de ejecuciones de Tests fueron realizadas por el Laboratorio UB consolidando los resultados con el fin de obtener el grado de Calidad asociado a cada Módulo.

Capacitación

Se realizaron cursos de capacitación en Calidad de Software y ESPECÍFICOS DE MyFEPs. Se pasó material de cursos con casos de estudio de los proyectos de evaluación MyFEPS a IRAM, y se dictaron dos talleres para SyK. A su vez se dictaron 2 cursos por el programa EMPLEARTEC, durante 2014 y 2015. Asimismo se incluyó MyFEPS en programas analíticos de asignaturas de carreras de Ingeniería Informática, Licenciatura en Sistemas.

Líneas de Investigación, Desarrollo e Innovación

Las líneas de investigación sobre las que se ha trabajado abarcan:

Dominio de conceptos: se ha definido un Glosario de términos relacionados a la calidad del software.

Herramientas de Apoyo a la Evaluación: se han investigado cuales son los Casos de Uso más Importantes que influyen en el proceso de evaluación y cómo afectan al proceso. Ejemplos de estos Casos son: a. Definición de los Objetivos de la evaluación de manera colaborativa, b. Definición del Modelo de Calidad a usar, c. Etapas y actividades del Proyecto, d. Ponderación e Integración de Resultados.

Intercambiabilidad del Modelo: se ha utilizado el Proceso MyFEPs, pero con los Modelos de calidad ISO/IEC 9126 e, ISO/IEC 25010, este último con la restricción de usar en lo posible las métricas ISO/IEC 9216-2 dado que a la fecha no se han publicado las métricas externas de la serie ISO/IEC 25000.

Método de evaluación de la calidad del software: se ha establecido un método de evaluación basado en los procesos de evaluación de software ISO/IEC 14598 e ISO/IEC 25040, pero incorporando actividades en la evaluación que permitiesen incorporar objetivos de evaluación de diferentes *stakeholders*, establecer mecanismos de ponderación de atributos a la hora de calcular la calidad del producto, e incorporando por cada atributo de calidad, un pseudo-procedimiento que indica cómo calcular sus métricas.

Si bien se ha validado que el método de evaluación MyFEPS es adecuado para realizar evaluaciones de calidad (incluso por estudiantes), y se han probado como

válidas algunas métricas, como ser aquellas que aplican a la evaluación de las características básicas “Satisfacción subjetiva de los usuarios”, “satisfacción de stakeholders no usuarios”, Usabilidad Objetiva”, “Facilidad de Instalación”, a modo de ejemplo, se considera apropiado continuar esta línea de investigación con evaluaciones que permitan validar la adecuación de las restantes características básicas (en total son 16) [3]

Resultados y Objetivos

El objetivo principal de los últimos años fue la continua experimentación y uso de MyFEPS (tanto con el modelo de calidad ISO/IEC 25000 como con QSAT) en diferentes tipos de proyectos académicos, (tesis) y profesionales (industria).

En ese sentido estuvo presente la transferencia a Organismos (IRAM, Gobierno, Universidades) y/o Industria (Laboratorios y Empresas) del Framework MyFEPS, el modelo QSAT, los avances en las normas ISO/IEC y los conocimientos prácticos capitalizados en los proyectos, en el uso de estos modelos.

Como Objetivos derivados de la continua experimentación, tuvimos:

1. Continuar validando la base teórica que de sustento al modelo QSAT y al proceso MyFEPS.
2. Mantener actualizado el acervo de metodologías, técnicas, documentación, herramientas y normas de evaluación de productos de software.
3. Experimentar en el uso del modelo de calidad, y la ponderación de características de calidad en base a objetivos de evaluación.

4. Especificar las métricas a utilizar para Medir el grado de calidad de cada característica fundamental.

5. Usar MyFEPS en diferentes escenarios, tanto metodológicos (para evaluar su adaptabilidad), como del tipo evaluación, y las particularidades de los softwares para cada tipo de producto y objetivos de Evaluación.

6. Usar MyFEPS para definir diferentes tipos de evaluaciones, sus lineamientos, técnicas, herramientas y plantillas para las actividades de evaluación.

7. Se validaron con proyectos piloto la validez del uso de MyFEPS, quedando QSAT desplazada por el modelo ISO 25023

8. A partir de los hallazgos encontrados en los pilotos, se validaron la pertinencia de los ajustes realizados a MyFEPS.

9. Se transfirieron los conocimientos adquiridos a la industria del software a través de conferencias, publicaciones y cursos de capacitación.

10. Se capitalizó los conocimientos (y se lo continua haciendo) adquiridos incorporándolos en asignaturas de las carreras de grado y en cursos de postgrado y de actualización profesional.

Formación de Recursos Humanos

Directora: Mg. Paula María Angeleri (2010-actual). Co Director Dr. Amos Sorgen (2010 al 2012).

Profesores investigadores: Mg. Rolando Titiosky (2011-actual), Lic. Jaquelina Wuille Bille (2011-2013).

Tesistas: Martín Santi, Agustín Ventura, Diego Ardizzone (tesinas de grado finalizadas), Roberto Basile, Facundo Scipione (tesinas de grado en curso).

Los integrantes de este grupo de investigación dirigen a estudiantes en diferentes contextos (becarios, tesinas, trabajos prácticos de cátedra). Además participan en el dictado de asignaturas de grado y cursos en los cuales este año se enseña el Framework MyFEPS.

Referencias

- [1] Sorgen, A., Angeleri, P.: El Modelo de Evaluación del Proyecto MyFEPS, 40JAIIO - ASSE 2011 - ISSN: 1850-2792, 180-191, 2011, <http://www.40jaiio.org.ar/sites/default/files/T2011/ASSE/790.pdf>
- [2] Angeleri, P.: Premios Sadosky 2008, IRAM Newsletter, Año 2 Nro.6, 2009.
- [3] Angeleri, P.; Oliveros, A; Sorgen, A.; Titiosky, R.; Wuille Bille J.; Modelo de calidad de productos de software, CONAISI 2014, ISSN: 2346-9927 - Página 1043.
- [4] Angeleri, P.; Titiosky, R.; Santi, M.; Davila, A.: Proceso de Evaluación de Productos *Software*, CONAISI 2015.
- [5] Sorgen, A., Angeleri, P.: Teoría y Práctica en la Evaluación de Productos de Software, XIV Workshop de Investigadores en Ciencias de la Computación 2012.
- [6] Titiosky, R., Angeleri, P., Sorgen, A., Wuille Bille, J.: "Proyectos de Evaluación de Productos de Software con un nuevo Framework de Calidad", XV Workshop de Investigadores en Ciencias de la Computación 2013, ISBN: 9789872817961.
- [7] Angeleri, P., Titiosky, R., Sorgen, A., Wuille Bille, J., Oliveros, A.: Ajustes al Framework de Evaluación de Productos de Software MyFEPS, XVI Workshop de Investigadores en Ciencias de la Computación 2014.
- [8] Angeleri, P., Sorgen, A., Bidone, P., Fava, A., Grasso, W.; Diseño y desarrollo de un framework metodológico e instrumental para asistir a la evaluación de software, 43 JAIIO - JUI 2014 - ISSN: 1851-2518 - Página 10.

http://sedici.unlp.edu.ar/bitstream/handle/10915/42015/Documento_completo.pdf?sequence=1

- [9] Ceballos, J.: Caso de Éxito de certificación de producto software según las normas serie ISO/IEC 25000, NEWSLETTER IRAM AÑO 9 • Nº 72 • Diciembre 2015.
- [10] Santi, M., Medición de la Usabilidad de un Ecommerce empleando el Framework MyFEPS, tesis de grado de Universidad de Belgrano, 2013.
- [11] Santi, M., Evaluación de la Usabilidad de un sitio de Comercio Electrónico empleando el Framework MyFEPS, CONAISI 2014.
- [12] serie de normas IRAM-ISO-IEC 14598: Tecnología de la información. Ingeniería de software. Evaluación del producto de software. Information technology. Software engineering. Software product evaluation. Parte 1, 2006, Partes 2,3, 2007, Partes 4,5, 2009, Parte 6, 2012.
- [13] ISO/IEC 25040 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Evaluation process. 2012
- [14] ISO/IEC 25010 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality model.
- [15] ISO/IEC 25000 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) – Overview
- [16] IRAM-NM-ISO/IEC 9126-1 Tecnología de la información-Ingeniería de software - Calidad del producto Parte 1 - Modelo de calidad (ISO/IEC 9126-1:2001, IDT) Information technology - Software engineering – Product quality Part 1 - Quality Model 1, IRAM, 2009.

IMPLEMENTACIÓN DE SISTEMAS INTELIGENTES PARA LA ASISTENCIA A ALUMNOS Y DOCENTES DE LA CARRERA DE INGENIERÍA EN SISTEMAS DE INFORMACIÓN

Pollo-Cattaneo, Ma. F.; Pytel, P.; Vegega, C.; Ramón, H.; Deroche, A.;
Straccia, L.; Bernal Tomadoni, L. & Acosta, M.

Grupo de Estudio en Metodologías de Ingeniería de Software. Facultad Regional
Buenos Aires. Universidad Tecnológica Nacional. Argentina.
{flo.pollo, ppytel}@gmail.com

Resumen

Considerando las particularidades de la asignatura 'Sistemas y Organizaciones' de la carrera de 'Ingeniería en Sistemas de Información' se busca desarrollar e implementar un Sistema Inteligente que asista el proceso de enseñanza-aprendizaje entre los alumnos y sus docentes. Para la construcción de dicho sistema informático se prevé la aplicación de tecnologías no convencionales provenientes de la Inteligencia Artificial, y, como resultado, se espera generar un beneficio a los alumnos, docentes y directivos de la carrera; así como contribuir con nuevo conocimiento a la comunidad académica.

Palabras clave:

Tecnología de la Educación, Medios Educativos, Inteligencia Artificial, Sistemas Inteligentes, Ingeniería de Software.

Contexto

El trabajo propuesto se desarrolla en el marco de actividades del Grupo de Estudio de Metodologías para Ingeniería en Software y Sistemas de Información (GEMIS), el cual se encuentra conformado por un equipo de docentes y alumnos dentro del ámbito de la Facultad Regional Buenos Aires de la Universidad

Tecnológica Nacional (UTN-FRBA). Este grupo busca la sistematización de cuerpos de conocimientos y promoción sobre el campo de la Ingeniería en Sistemas de Información y la Ingeniería en Software, sus aplicaciones y abordajes metodológicos en todo tipo de escenarios (convencionales y no convencionales).

Introducción

De acuerdo al Consejo Federal de Decanos de Ingeniería (CONFEDI), sostiene que en países desarrollados es aceptable que por año se gradúe 1 nuevo ingeniero cada 2.000 a 2.500 habitantes; sin embargo en Argentina, esta tasa se encuentra en el orden de 1 nuevo ingeniero cada 7.600 habitantes [1]. La consecuente falta de ingenieros genera que, muchas empresas presenten tentadoras ofertas de trabajo y reclutamiento de pasantes, anticipándose a la graduación de estudiantes avanzados [2]. Sin embargo, a pesar de que esta situación fomenta la captación de nuevos aspirantes, genera una alta deserción. Por consiguiente, y en el marco del Plan Estratégico de Formación de Ingenieros (PEFI) 2012–2016 [3], las universidades argentinas apuntan a llegar en 2016 a los 10 mil egresados por año, es decir, a 1 ingeniero cada 4.000 habitantes. A tal efecto, se destacan los sistemas de tutoría implementados en los primeros años de

las carreras por muchas universidades con el objetivo de compensar el déficit de la educación media [4]. Sin embargo, se considera que todavía queda trabajo por realizar para cumplir con las demandas profesionales del sector empresarial, la ampliación de la masa de aspirantes a ingenieros en las aulas y la concreción de la finalización de una formación académica adecuada.

Dentro del marco del PEFI se ha definido como una de las acciones a desarrollar la aplicación de herramientas informáticas para la enseñanza y el aprendizaje [3]. Como consecuencia, tanto en universidades públicas y privadas del país se fomentó el uso de aulas virtuales y se dictaron capacitaciones asociadas al uso de las Tecnologías de la Información y la Comunicación (TIC). Esto ha motivado a que la comunidad educativa haya encontrado en las TIC una herramienta muy útil para asistir el proceso de enseñanza-aprendizaje, sobre todo con respecto a la facilidad en el manejo de los recursos asociados a los contenidos [5]. De estas herramientas se destaca MOODLE (Modular Object-Oriented Dynamic Learning Environment) como uno de los entornos de aprendizaje más usados [6].

No obstante, el solo hecho de incluir una herramienta informática en el aula no significa que se generen aportes positivos a la construcción del conocimiento [7]. Es imprescindible que la herramienta se encuentre adaptada a las necesidades y posibilidades de los estudiantes y docentes. Esto significa que, además de facilitar el acceso al contenido de recursos en cantidad y calidad, debe promover la reflexión y el debate en los procesos de comunicación tanto dentro como fuera del aula [8]. De esta manera, el estudiante puede adquirir habilidades cognitivas e instrumentales, así como entrenarse en la experiencia de aprender a aprender o, del

trabajo en equipo, entre otros aspectos, contribuyendo así a una actividad claramente proactiva conformada por un ambiente de aprendizaje colaborativo y una cultura basada en el compartir contenidos y significados de esos contenidos. Por lo tanto, e-Learning no alcanza su potencial simplemente al colocar en una herramienta guías, recursos, preguntas frecuentes, consejos técnicos, enlaces a páginas web, acceso a bases de datos y/o bibliotecas virtuales para que sean accesibles a los alumnos, sino que se requiere también de un conjunto de servicios de apoyo para que el estudiante pueda interactuar con los contenidos y pueda sentirse motivado para cumplir con sus objetivos [9]. Las actividades antes mencionadas, requieren la presencia de un tutor que complemente el trabajo del docente en clase, mediante el acompañamiento del alumnado en un contexto de aprendizaje activo. Entre sus principales tareas se encuentra la monitorización del aprendizaje y la evaluación continua junto con un apoyo técnico y pedagógico permanente [10].

La automatización de estas tareas requiere el diseño de sistemas informáticos realmente adaptativos e inteligentes capaces de adecuarse de manera autónoma y rápidamente a la multiplicidad de variables que forman parte de un contexto formativo. Sin embargo, estos requerimientos no pueden ser alcanzados a través del uso de software convencional, es necesario aplicar principios y técnicas de la Inteligencia Artificial [11]. Como resultado, nace el concepto de ‘Sistema Tutorial Inteligente’ (STI) para describir la tecnología que pretende emular el comportamiento de un tutor humano en un entorno de e-learning [12]. Para ello, tiene la capacidad de asistir al proceso de enseñanza-aprendizaje humano, adaptando el tipo y el contenido de la

instrucción a las necesidades específicas de cada alumno. Esto significa que, un STI debe analizar el comportamiento de cada alumno tanto para conocer su estado de conocimiento como para satisfacer sus necesidades mediante la selección y aplicación de las estrategias de enseñanza más adecuadas [13]. Éstas deben resolver una serie de cuestiones prioritarias: qué explicar, con qué nivel de detalle, y cómo detectar y corregir los errores de los alumnos. Por ello, a diferencia de las aplicaciones de e-learning basadas en tecnología web, actúan como entrenadores, ofreciendo sugerencias cuando los estudiantes dudan o se atascan en el proceso de la resolución de un problema, y no sólo cuando ellos introducen la respuesta [14]. Finalmente, permiten hacer un mejor uso de las posibilidades tecnológicas brindando, además de la información solicitada, los mecanismos que ayuden a convertir dicha información en conocimiento [15].

Líneas de Investigación, Desarrollo e Innovación

A partir de la aplicación de diferentes mecanismos tecnológicos dentro del ámbito educativo universitario de la carrera de 'Ingeniería en Sistemas de Información' de la UTN-FRBA, se han detectado diversos problemas de aprendizaje en los alumnos de la asignatura 'Sistemas y Organizaciones' [16]. Dicha asignatura corresponde al primer nivel de la carrera y es de cursado obligatorio para los alumnos que hayan aprobado el curso de ingreso a la carrera (aproximadamente 800 inscriptos). Por consecuencia, cualquier tipo de intervención llevada a cabo en los cursos de esta asignatura garantiza un impacto sobre la totalidad de la población ingresante.

Entre los problemas detectados, los cuales han sido documentados en [17-20], se destaca el hecho que muchos alumnos desaprueban el primer examen parcial de la asignatura. En algunos casos, ello produce una frustración tal que deriva en la interrupción de la cursada para su retome al año siguiente (generando un retraso significativo en la finalización de la carrera) o, en casos más drásticos, el abandono de la carrera (aumentando así la deserción universitaria).

En este contexto, se ha decidido llevar a cabo la implementación de un Sistema Inteligente que incluya un conjunto de funcionalidades para dar solución a los problemas específicos de la asignatura. Dicho sistema será la semilla para un futuro STI que podrá llegar, mediante la incorporación de mejoras y nuevos módulos, a su arquitectura.

En su versión inicial se prevé el desarrollo de las siguientes funcionalidades principales:

- (a) *Evaluación del conocimiento en los temas teóricos de los alumnos mediante tecnología móvil.* Es decir, generar estrategias de evaluación de los conceptos teóricos mediante la capitalización de los dispositivos móviles en la población estudiantil con la posibilidad de m-learning [21]. De esta manera, en todo lugar y aprovechando los tiempos ociosos, los alumnos podrán valorar su conocimiento respondiendo preguntas del estilo de opción múltiple idóneas a la asignatura.
- (b) *Revisión automática de los diagramas de organigramas y cursogramas* realizados por los alumnos como parte de la ejercitación práctica. Para ello, el módulo hará uso de tecnologías correspondientes a la Visión Artificial [22] que permitirán el procesamiento automático de los ejercicios

realizados. Una vez analizadas las características de los diagramas suministrados, y realizando una revisión contra el enunciado del ejercicio, este módulo le ofrecerá al alumno una valoración cuantitativa del ejercicio así como un conjunto de observaciones cualitativas sobre correcciones o mejoras que se podrían incluir.

- (c) *Análisis del nivel de conocimiento de los alumnos sobre los temas teóricos y prácticos de la asignatura.* A partir de la información recolectada por los módulos (a) y (b), y, aplicando mecanismos del Razonamiento Aproximado [23] este módulo podrá determinar el nivel de conocimiento que tiene cada alumno. De esta manera se buscan identificar las fortalezas y debilidades de los estudiantes, información que será utilizada por los módulos (d) y (e) que son descriptos a continuación.
- (d) *Generación de recomendaciones a los alumnos* sobre los temas en cuyo estudio deben profundizar. Al combinar la información generada por (c) con un conjunto de ontologías previamente definidas a partir de la experiencia de los docentes, este módulo tendrá la capacidad de identificar los medios (apuntes, bibliografía, ejercicios propuestos, etc.) que deberá utilizar el alumno para mejorar sus conocimientos. Con el objeto de implementar una interface sencilla, basada en el Procesamiento del Lenguaje Natural [24], e incorporará principios de los Sistemas Multi-Agentes [25].
- (e) *Generación de reportes sobre los alumnos.* Utilizando toda la información recolectada, este módulo suministrará información relevante y oportuna, tanto a nivel general (por

curso) o particular (por alumno), para detectar posibles problemas de aprendizaje y poder seleccionar las estrategias didácticas para darles solución.

Resultados y Objetivos

El proyecto propuesto se inscribe en una línea de investigación dentro del ámbito educativo universitario mediante la aplicación de tecnologías no convencionales. De esta manera, se busca continuar y ampliar los trabajos desarrollados por GEMIS en el ámbito educativo implementando artefactos tecnológicos que permitan asistir a docentes y alumnos en sus tareas tanto dentro como fuera del aula. Concretamente, se pretende suministrar a los alumnos una oportunidad para el repaso, el entrenamiento y la puesta a prueba de la profundidad que efectivamente se tiene sobre el manejo de los temas teóricos y prácticos aprendidos durante la cursada de la asignatura, antes del inevitable acto final de un examen. En forma complementaria, se generará información que podrá ser utilizada por los docentes para seleccionar y aplicar estrategias didácticas pertinentes en cada uno de sus cursos.

Formación de Recursos Humanos

El equipo de trabajo se encuentra conformado por tres investigadores formados, un tesista de maestría, dos graduados de grado, un investigador de apoyo y un alumno en la carrera de Ingeniería en Sistemas de Información. Asimismo, se prevé incorporar alumnos avanzados en la carrera de Ingeniería en Sistemas de Información con posibilidades de articular sus Trabajos Finales de Carrera de Grado, así como, también dos tesistas de posgrado que

desarrollarán sus Trabajos Finales Integrador de Especialidad y/o Tesis de Maestría. De esta manera se espera generar un verdadero espacio integrado de investigación en el área de Sistemas de Información en el nivel de carreras de grado y posgrado.

Referencias

- [1]. Consejo Federal de Decanos de Ingeniería (2010) La Formación del Ingeniero para el Desarrollo Sostenible - Aportes del CONFEDI .Congreso Mundial Ingeniería 2010 (Buenos Aires) <http://tinyurl.com/l35s8w9> Disponible online en Marzo de 2016.
- [2]. iProfesional (2011) Ingenieros: las empresas pagan sueldos muy altos, pero es una "misión imposible" encontrarlos. <http://tinyurl.com/lcixqzz> Disponible en Marzo de 2016.
- [3]. Secretaría de Políticas Universitarias (2012) Plan Estratégico de Formación de Ingenieros 2012-2016. Ministerio de Educación de la Nación Argentina <http://tinyurl.com/kz9lrpz> Disponible online en Marzo de 2016..
- [4]. Clarín.com (2015) Ingeniería resurge con más graduados y nuevas vocaciones. <http://tinyurl.com/oogvt97> Disponible online en Marzo de 2016.
- [5]. Betancur, D., Moreno, J. & Ovalle, D. A. (2009). Modelo para la recomendación y recuperación de objetos de aprendizaje en entornos virtuales de enseñanza/aprendizaje. Revista Avances en Sistemas e Informática, 6(1).
- [6]. Poveda, L. A. & de la Educación, D. F. (2007). Moodle como recurso didáctico. EDUTEC, 2-3.
- [7]. Salinas, J. (2004) Innovación docente y uso de las TIC en la enseñanza universitaria. En Revista de Universidad y Sociedad del Conocimiento. UniversitatOberta de Catalunya, España, pp. 1-16.
- [8]. GRIAL (2011) e-Learning in Spain. Technical Report. <http://tinyurl.com/n6z3qu4> Disponible online en Marzo de 2016.
- [9]. Seoane Pardo, A. M. & Zangrando, V. (2006) El factor humano en el eLearning: Tutor online. Profesiones emergentes: especialista en eLearning. Salamanca (ed. electrónica)
- [10]. Seoane Pardo, A. M. & García Peñalvo, F. J. (2006) Criterios de calidad en formación continua basada en eLearning. Una propuesta metodológica de tutoría online. <http://tinyurl.com/mb39dg4> Disponible online en Marzo de 2016.
- [11]. Russell, S. J., Norvig, P., Davis, E., Russell, S. J. y Russell, S. J. (2010) Artificial intelligence: a modern approach (Vol. 2). Englewood Cliffs: Prentice hall.
- [12]. de Arriaga Gómez, F. (2012). E-learning inteligente: un instrumento para la formación permanente. Doctoral dissertation, Universidad Nacional de Educación a Distancia. <http://tinyurl.com/o7pxvj2> Disponible online en Marzo de 2016.
- [13]. Ferrero, B., Arruarte, A., Fernández-Castro, I., & Urretavizcaya, M. (2001). Herramientas de autor para enseñanza y diagnóstico: IRIS-D. Inteligencia Artificial, 12, 13-28.
- [14]. Sánchez Vila, E. & Lama Penín, M. (2007). Técnicas de la Inteligencia Artificial aplicadas a la Educación. Revista Iberoamericana de Inteligencia Artificial, (33), 7-12.
- [15]. Rodríguez, C. L. R. R., Brito, C. M. L. & Piña, A. A. D. (2013) Agentes inteligentes para proporcionar adaptabilidad a un ambiente de enseñanza aprendizaje asistido por computadora. Revista Congreso Universidad Vol. 2, No. 1. ISSN: 2306-918X.
- [16]. UTN-FRBA (2008) Programa analítico de la cátedra 'Sistemas y Organizaciones' - Plan 2008. <http://tinyurl.com/oh6mrk4> Disponible online en Marzo de 2016.
- [17]. Raus, N. A., Lujan, F. N., Deroche, A., Vegega, C., Pytel, P., Pollo-Cattaneo, M. F. (2013) Aplicación del Proceso de Ponderación de Reglas de Pertenencia a Grupos en Evaluaciones Finales en Carreras de Grado. Memorias del 1er Congreso Nacional de Ingeniería Informática / Sistemas de Información (CoNaIISI 2013). ID 22-440-1-DR. ISSN: 2346-9927.
- [18]. Pollo-Cattaneo, M.F, Deroche, A., Raus, N., Lujan, F., Vegega, C., Pytel, P. (2013). Análisis de exámenes en carreras de Sistemas mediante procesos de Explotación de Información. En Reflexiones sobre Ingeniería de Requisitos y Pruebas de Software (Ed. Jaime Echeverri). Pág. 97-111. Editorial de la Corporación Universitaria Remington y Organización LACREST. ISBN 978-958-58070-3-7.
- [19]. Pollo-Cattaneo, M.F., Acosta, M., Straccia, L., Bernal, L., Lujan, F., Raus, N., Deroche, A., Vegega, C. & Pytel, P. (2014) Identificación de Perfiles de Alumnos y Análisis de sus Percepciones sobre la carrera de Ingeniería en Sistemas de Información mediante Explotación de Información. Workshop de Innovación en Educación en Informática. Artículo 6537. Proceedings XX Congreso Argentino de Ciencias de la Computación. ISBN 978-987- 3806-05-6.
- [20]. Acosta, M., Straccia, L., Bernal, L., Vegega, C., Pytel, P. & Pollo-Cattaneo, Ma. F. (2015) Avances en el Análisis del Seguimiento de Perfiles de Alumnos y de sus Percepciones sobre la carrera de Ingeniería en Sistemas de Información mediante Explotación de Información. Memorias de 3er Congreso Nacional de Ingeniería Informática y Sistemas de Información (CONAISI 2015). Workshop de Educación en Ingeniería. Artículo 49-587. ISBN 978-987-1896-47-9.
- [21]. Herrera, S. & Fennema, M. (2011) Tecnologías Móviles Aplicadas a la Educación Superior. XVII Congreso Argentino de Ciencias de la Computación, pp. 620-630 ISBN: 978-950-34-0756-1.
- [22]. Reinhard Klette (2014) Concise Computer Vision. Springer. ISBN 978-1-4471-6320-6.
- [23]. Barber, D. (2012) Bayesian Reasoning and Machine Learning. The MIT Press.
- [24]. Cortez Vásquez, A., Pariona Quispe, J., & Huayna, A. M. (2014). Procesamiento de Lenguaje Natural. Revista de investigación de Sistemas e Informática, 6(2), 45-54.
- [25]. Byrski, A., Drezewski, R., Siwik, L., & Kisiel-Dorohinicki, M. (2015). Evolutionary multi-agent systems. The Knowledge Engineering Review, 30(02), 171-186.

Líneas de Investigación de Realidad Aumentada

Aplicada a la Asistencia Médica en el campo de la Emergentología

Nicolás Verdicchio, Diego Sanz, Jonathan Barth, Cristian Montalvo, Facundo Petrollo, Nahuel Mangiarua, Santiago Igarza, Jorge Ierache

Universidad Nacional de La Matanza, Departamento de Ingeniería e Investigaciones Tecnológicas.

Grupo de Realidad Aumentada Aplicada

Florencio Varela 1903, La Matanza, Buenos Aires, Argentina
jierache@unlam.edu.ar

Resumen

Este trabajo presenta la descripción del desarrollo de las líneas de investigación aplicadas a la aumentación de aplicaciones para la salud, que tienen como fin mejorar el tiempo de asistencia a un individuo y facilitar información médica de carácter relevante en situaciones de emergencia en la vía pública. Estas aplicaciones están diseñadas para utilizarse en un dispositivo móvil, por lo que permite su uso tanto a la población civil como al personal médico.

Palabras clave: Salud, Realidad Aumentada, Primeros Auxilios, Historial Médico, Tarjeta médica.

Contexto

La investigación es desarrollada por el grupo de investigación de Realidad Aumentada Aplicada del Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza, en el marco del proyecto PROINCE C168. El grupo se encuentra financiado a través de las becas otorgadas por la institución.

Introducción

En primeros auxilios, atender al individuo de forma rápida y precisa resulta un elemento vital en situaciones de emergencias médicas. Para lograr esto, debemos obtener conocimiento del historial médico del afectado para saber de sus alergias, enfermedades pasadas o posibles operaciones que pueden afectar la ayuda al mismo. Existen aplicaciones como “My Medical Info” [1], “Medref” [2] con las que podemos mantener un registro de las enfermedades y posibles patologías que posee el paciente. La aplicación “Simple Medical Information” [3] es utilizada para obtener un diagnóstico prematuro dependiendo de los síntomas que señalemos.

La línea de investigación que desarrolla el grupo de realidad aumentada en el marco de la asistencia médica pretende sumar capacidades diferenciales considerando empleo de la tecnología de Realidad Aumentada (a partir de aquí denominada RA) con el fin de contribuir con la asistencia médica con el objetivo de mejorar el tiempo de respuesta ante una emergencia médica u accidente de una persona.

La RA permite la fusión de los datos virtuales y objetos sobre el mundo físico, enriqueciendo la percepción de la realidad, aumentándola [4]. Se suele comparar esta tecnología con Realidad Virtual, pero no representan el mismo concepto. Realidad Virtual implica la inmersión en un ambiente completamente virtual mientras que RA consiste en mantenerse en el mundo real mientras se despliegan y muestran contenidos virtuales superpuestos sobre él. [5].

Líneas de Investigación, Desarrollo e Innovación

La línea de investigación aplicada se centra en el diseño de un prototipo de sistema de tarjeta aumentado para el servicio de emergencias médicas que denominaremos por sus siglas en inglés ARCSEMS. (Augmented Reality Card System for Emergency Medical Services) Es una herramienta que permite la gestión y explotación de datos en dispositivos móviles que utilizan RA, el cual presenta virtualmente información médica relevante, al aumentar un objeto físico clásico de un individuo (ficha, credencial, chapa de identificación colgante, etc.), al momento de asistirlo frente a una situación de emergencia, proporcionando datos relevantes para la persona que lo asiste, ayudándole así a tomar una decisión sobre el curso de acción que debe de tomar para el tratamiento exitoso del afectado.

Se complementa la línea de investigación con la incorporación de sistemas basados en conocimiento con la finalidad de alcanzar una clasificación del individuo en función de sus antecedentes médicos.

Resultados y Objetivos

En el marco de los resultados de línea de investigación de sistema de catálogo virtual aumentado [6], [7], [8], y la idea de contribuir al desarrollo de aplicaciones que solucionen problemas referidos al ámbito de la salud y que estén destinadas a mejorar la calidad de vida de la sociedad, surge como resultado esta línea de investigación, que es la implementación de la primera versión del Sistema de Tarjetas Virtuales Aumentadas [9] el cual consiste en el despliegue de datos de relevancia para el usuario sobre una tarjeta, mediante el enfoque de un "Marker" (Imagen impresa que proporciona una referencia espacial, permitiendo al dispositivo móvil superponer la información virtual en el entorno real captado, conocido como "Marker" que se utiliza como punto de referencia para posicionar la información).

Con el fin de mantener organizados los datos de cada persona en una ficha médica que contiene datos personales (nombre, apellido, grupo sanguíneo, números de teléfonos de parientes cercanos), datos médicos (enfermedades padecidas anteriormente, operaciones que le fueron realizadas, alergias, etc.) y datos de carácter complementario (Si es donador de órganos, si es fumador o alguna condición especial que no recaiga en los apartados anteriores).

El desarrollo del prototipo se hace mediante una Web App preliminar desarrollada específicamente, sobre la cual se pueden revisar los datos médicos de las personas y editarlos, en el caso de que necesiten hacerlo.

El módulo móvil incluye una aplicación compilada para dispositivos móviles con

sistema operativo Android que utilizan el sistema multimedia Unity3D motor y el reconocimiento Vuforia.

La Figura 1 muestra la aplicación móvil y su iteración con el Frame Marker (dispositivo móvil y el reconocimiento del marcador en este caso sobre una clásica chapa de identificación colgante). En la figura 2 se muestra la aumentación de información del usuario que se está asistiendo, el contenido aumentado se puede visualizar en la pantalla del dispositivo que en este caso se compone de un conjunto de botones de color rojo en la parte superior. De izquierda a derecha, se representa las opciones de: a) llamada de emergencia (911), b) exhibición de contenido aumentado de datos médicos, c) contactos personales (familiares), d) la opción de activación permanente para la visualización de contenido de información independiente del foco sobre marcador. Debajo de los botones podemos ver el área de edición de los datos médicos pertinentes de la persona.

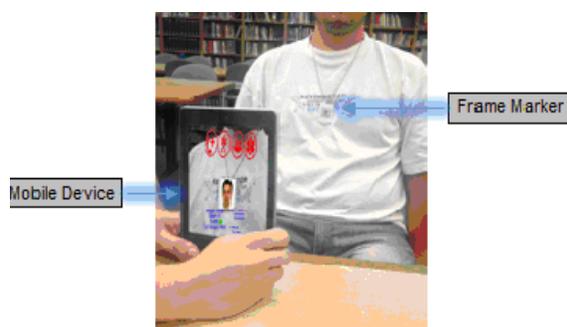


Fig. 1 Imagen del ARCSEMS aumentando el colgante identificador



Fig. 2 Contenidos aumentados en el dispositivo móvil

En la actualidad, ante una situación de emergencia, los centros hospitalarios suelen emplear sistemas de clasificación o categorización de la atención de urgencia, con el fin de acelerar y optimizar la atención brindada y poder asistir o derivar a cada paciente correctamente. Pero, en su mayoría, estos sistemas presentan la falencia de tener que trabajar “a ciegas” con los pacientes, por lo que los factores a tener en cuenta en dicha categorización se ven limitados. Es aquí donde se presenta la necesidad de incorporar a las funcionalidades del Sistema de Tarjetas Virtuales Aumentadas, un Sistema basado en conocimientos que logre categorizar al paciente en función de sus antecedentes y de esta forma contribuir a definir un estado del paciente a partir de su historial médico ya incorporado al mismo. Este estado médico del paciente brindará una información de forma rápida a la persona que lo esté asistiendo, la cual podrá identificar visualmente la clasificación del paciente en función de sus antecedentes.

El Sistema basado en conocimientos incluso podrá beneficiarse de datos médicos en tiempo real con el uso de los dispositivos biométricos y así poder brindar un estado del paciente más acertado. En el marco de las futuras líneas de investigación se considera la obtención de datos en tiempo real

mediante el uso de wearables, los cuales incluyen varios sensores biométricos como sensores de pulsaciones y de ritmo cardíaco los cuales se utilizarán para actualizar la base de conocimientos del paciente en forma constante y saber en qué situación se encuentra el mismo y lograr enviar una alerta automática a emergencias en el caso de que el ritmo cardíaco aumente considerablemente o el pulso disminuya para que éstos puedan atender a la situación de forma rápida.

Formación de Recursos Humanos

El grupo de investigación se encuentra conformado por un investigador formado, dos investigadores en formación, dos graduados, y cinco alumnos del departamento de Ingeniería e Investigaciones Tecnológicas de la UNLaM, trabajando en el área de Realidad Aumentada. La formación de los integrantes tanto en el área técnica como en el mecanismo formal de investigación y publicación conforma una importante porción de nuestro trabajo. Adicionalmente el grupo participa en los distintos eventos científicos tecnológicos abiertos a la comunidad que organiza la universidad tales como Expo Proyecto y la Bienal de Arte, como así también en los congresos de la RedUNCI con el fin de acercar el conocimiento obtenido y generado al público general.

Referencias

1. My Medical Info. Retrieved November, 2015, from: <https://play.google.com/store/apps/details?id=com.murryelectronics.MyMedicalInfo&hl=es>
2. MedRef for Glass. Retrieved November, 2015, from: <https://medrefglass.appspot.com/>
3. Simple Medical Information. Retrieved November, 2015, from: <https://play.google.com/store/apps/details?id=com.simplemedicalinformation.smi>
4. Manresa Yee, M. Abásolo, R Más Sansó and M Vénere (2011). Realidad virtual y realidad aumentada. Interfaces avanzadas. ISBN 978-950-34-0765-3
5. Cristina Manresa Yee, María José Abásolo, Ramón Más Sansó and Marcelo Vénere (2011). Realidad virtual y realidad aumentada. Interfaces avanzadas. ISBN 978-950-34-0765-3
6. Bevacqua, Sebastián Ariel; Igarza, Santiago; Mangiarua, Nahuel Adiel; Becerra, Martín Ezequiel; Verdicchio, Nicolás Nazareno; Ortiz, Fernando Martín; Sanz, Diego Rubén; Duarte, Nicolás Daniel; Sena, Matías Ezequiel; Ierache, Jorge Salvador “Líneas de Investigación del Grupo de Realidad Aumentada Aplicada de UNLaM Catálogos Virtuales Aumentados - Aumentación de Material Didáctico”, XVII WIIC, Salta 2015., ISBN 978-987-633-134-0.
7. Ierache J., Mangiarua N., Bevacqua S., Becerra M., Verdicchio N., Duarte N., Sanz D., Ortiz F., Sena M., Igarza S., “Sistema de Catálogo para la Asistencia a la Creación, Publicación, Gestión y Explotación de Contenidos Multimedia y Aplicaciones de Realidad Aumentada”, XX CACIC 2014., San Justo La Matanza. ISBN 978-987-3806-05-6

8. Ierache J., Mangiarua N., Bevacqua S., Becerra M., Verdicchio N., Duarte N., Sanz D., Ortiz F., Sena M., Igarza S “Development of a Catalogs System for Augmented Reality Applications”, Singapur 2015, World Academy of Science, Engineering and Technology (WASET).Internacional Science Index ISSN 1307:6892
9. Jorge Ierache, Nicolás Verdicchio, Nicolás Duarte, Cristian Montalvo, Facundo Petrolo, Diego Sanz, Jonathan Barth, Nahuel Mangiarua, Santiago Igarza, “Augmented Reality Card System for Emergency Medical Services”, IWBBIO 2016 (4th International Work-Conference on Bioinformatics and Biomedical Engineering) 20 - 22 abril 2016, 8 pp en prensa.

Informatización de datos Botánicos de la Universidad Nacional de Luján: un camino al Sistema de Bioinformática de la Institución

Panessi Walter, Ortiz Claudia, Apóstolo Nancy, Perroud Claudia

Departamento de Ciencias Básicas, Universidad Nacional de Luján (UNLu)
wpanessi@unlu.edu.ar, cortiz@unlu.edu.ar, nancy.apostolo2@gmail.com,
cperroud@unlu.edu.ar

Resumen

Desde hace unos años los docentes e investigadores de Botánica de la Universidad Nacional de Luján han realizado diversos viajes de campo y trabajos de investigación que han provisto abundante información de especies nativas e introducidas de nuestro país. Asimismo, se han iniciado proyectos y gestiones para la Creación del Jardín Botánico de la institución. Este proyecto tiene como objetivo estudiar las posibles metodologías de informatización de datos botánicos para asociarlos a sistemas de bioinformática destinados a diversos usuarios de la Comunidad Universitaria y General, iniciando el proceso de bioinformatización institucional de la Universidad Nacional de Luján. Como producto de este plan de trabajo se espera lograr una base de datos integrada con los materiales fotográficos y textuales producidos hasta el momento y con la constante provisión de nuevos aportes derivados de futuros trabajos y viajes de campo. Dicha base será de acceso a docentes e investigadores de la Casa de Estudios, así como a público general. Además, los contenidos de la misma podrán ser vinculados a los ejemplares del Jardín Botánico de la UNLu, permitiendo a los

visitantes acceder a información específica de cada especie.

Palabras clave: Bioinformatización, Software multimedial, Software adaptado al contexto

Contexto

Este trabajo de investigación se encuentra radicado en la Universidad Nacional de Luján, en su Departamento de Ciencias Básicas. En el marco de la convocatoria a proyectos de investigación y/o desarrollo tecnológico interdisciplinarios. El inicio fue en 2014 y tiene como duración 4 años (2014-2018). En el mismo trabajan docentes investigadores de la División Biología, área Botánica y de la División Computación, áreas Ingeniería de Software, Bases de Datos y Sistemas de Información y el área Algoritmos y Lenguajes. Participan también alumnos de las carreras Licenciatura en Sistemas de Información, Licenciatura en Ciencias Biológicas e Ingeniería Agronómica de la Universidad Nacional de Luján.

Introducción

La conservación de la biodiversidad exige constantemente la toma de decisiones importantes con implicaciones sociales,

políticas y económicas. Por ello, demanda un acceso fácil a información relevante y de alta calidad científica para que los usuarios y tomadores de decisiones dispongan con celeridad de los datos necesarios para conservar organismos. Las tecnologías de información son una herramienta coyuntural que permitirían enfrentar este reto (Font, 2005; Mata Montero & Mata, 2006).

La bioinformática es una nueva área interdisciplinaria que desarrolla y utiliza tecnología de información y de comunicación (TIC) en los procesos de generación, procesamiento y divulgación de información para apoyar la conservación de la biodiversidad. Para ello, deben articularse un proceso con tres tareas básicas: captura de información y conocimiento, análisis e interpretación de los datos para convertirlos en información más elaborada y, finalmente, la transferencia de la información procesada para distintos tipos de usuarios mediante diversos formatos.

Las TICs cumplen un rol importante en la conservación de la biodiversidad (Mata Montero & Mata, 2006).

La implementación de sistemas bioinformáticos en una Institución debe pasar por una serie de fases, tal como lo han indicado investigadores del Instituto Nacional de Biodiversidad (INBio) de Costa Rica. La fase inicial es el procesamiento manual de los datos mediante notas e informes de campo y laboratorio, información asociada a colecciones de especímenes, etc., los cuales se mantienen en sistemas analógicos como materiales impresos, fichas, dibujos, imágenes digitales, fotografías, videos, entre otros. En una

segunda fase, los científicos generan y diseñan, en forma independiente y sin integrarlo al resto de la institución, sus propias bases de datos mediante el uso de variados programas. La siguiente fase se relaciona a los sistemas de información integrados institucionalmente. En esta etapa comienzan a intervenir los profesionales informáticos que realizan los procesos básicos de diseño, programación y mantenimiento de los programas de acuerdo a las necesidades institucionales e implementando un proceso de migración de datos de las bases propias de los científicos hacia la base de datos institucional. A partir de aquí, se efectúan otras acciones que conllevan a sistemas de información institucional interdisciplinaria entre las diferentes áreas de la Biología y sistemas de información interinstitucionales, provinciales y/o nacionales (Mata Montero & Mata, 2006). Este proceso, actualmente, está ejemplificado no solo en el INBio, sino también en el proyecto mundial GBIF (Global Biodiversity Information Facility), en el proyecto europeo BIOCASE (A Biological Collection Access Service for Europe) y en escala regional española BDBC (Banco de datos de biodiversidad de Cataluña). Se han emprendido varios esfuerzos regionales y nacionales para el desarrollo de informatización de datos botánicos (Beltrán, 2005; Carreras et al., 2008; Cerana et al., 2007; Cosa et al., 2011, 2012, 2013; Dottori et al., 2008; Mariño et al., 1998) y de Herbario (FCNyM, 2014; IBODA, 2014; IMBIV, 2014; MACN, 2014). La mayoría de estas colecciones informatizadas forman parte del Sistema Nacional de Datos Biológicos impulsado por el Ministerio Nacional de Ciencia,

Tecnología e Innovación Productiva. El SNDB tiene como misión conformar una base de datos unificada de información biológica de uso y de catálogos sobre recursos naturales y otros temas afines. Entre sus objetivos se menciona promover el intercambio de información biótica a través de una red nacional de datos, así como analizar y acordar políticas conjuntas sobre calidad y distribución de datos y dotar de proyección internacional a los datos biológicos producidos en el país a través de su difusión en redes virtuales. (SNDB, 2014).

Desde 2009, el grupo de docentes e investigadores de Botánica del Departamento de Ciencias Básicas de la Universidad hemos emprendido una serie de viajes de observación de flora in situ en diversos puntos del país. En cada viaje se programó la participación de docentes o investigadores conocedores de la flora del lugar a visitar con el fin de lograr una cabal identificación de las especies observadas y un intercambio de conocimientos con dichos profesionales. Producto de ello, se ha obtenido una abundante información basada en imágenes captadas de especies identificadas en cada sitio visitado. Asimismo, se han colectado ejemplares de herbario mediante los permisos correspondientes en algunos sitios y se han adquirido plantas en maceta de especies nativas en viveros especializados o gubernamentales. Dichas plantas serán cultivadas en el predio destinado al Jardín Botánico de la UNLu como parte de las tareas programadas en un proyecto de extensión provisto por el grupo de docentes de Botánica y aprobado por

Disposición CDD-CB N° 100/13 y Resolución HCS N° 196/13.

Toda esta información botánica recabada durante varios años es necesaria para cumplir con las dos primeras fases del proceso de implementación de un sistema bioinformático institucional. Para continuar con las siguientes etapas de las fases del procedimiento es de trascendental valor la participación de profesionales informáticos para lograr el trabajo interdisciplinario fundamental

Líneas de Investigación, Desarrollo e Innovación

La documentación que el grupo de docentes ha registrado está disponible digitalmente y almacenado en un conjunto de archivos. Estos archivos fueron separados en un árbol de carpetas (o directorios) para formar con todo ello una base de datos de referencia que el grupo de Botánica utiliza y que tiene a su disposición en una computadora.

Las limitaciones impuestas por la poca flexibilidad que brinda un conjunto de archivos en una máquina estanca, en conjunto con la pretensión de vincular a los visitantes del próximo Jardín Botánico de la UNLu y esta base de datos, seguramente ampliada, motivan el impulso de generar un sistema informático multimedial y disponerlo en un servidor de datos que sea accesible públicamente. Por otra parte, para que esta información sea útil a la comunidad de usuarios, también se requiere un conjunto de aplicaciones software que permitan la consulta y la recuperación de la información y su distribución en el formato más adecuado, probablemente utilizando Information Retrieval como método de recuperación de información.

Para lograr la vinculación mencionada se requerirá el desarrollo de interfaces que permitan la recuperación de información sensible al contexto del visitante.

A pesar de que en Argentina se realizan abundantes emprendimientos de informatización de datos biológicos, en la UNLu no se han iniciado tareas de bioinformática.

Resultados y Objetivos

Como producto de este trabajo se espera lograr una base de datos integrada con los materiales fotográficos y textuales producidos hasta el momento y con la constante provisión de nuevos aportes derivados de futuros trabajos y viajes de campo. Dicha base será de acceso a docentes e investigadores de la Casa de Estudios, así como a público general. Asimismo, los contenidos de la misma podrán ser vinculados a los ejemplares del Jardín Botánico de la UNLu, permitiendo a los visitantes acceder a información específica de cada especie.

Estudiar las posibles metodologías de informatización de datos botánicos para asociarlos a sistemas de bioinformática destinados a diversos usuarios de la Comunidad Universitaria y General.

Iniciar el proceso de bioinformatización institucional de la Universidad Nacional de Luján.

Formación de Recursos Humanos

El proyecto está actualmente integrado por: cuatro profesores adjuntos, cinco docentes auxiliares, de las divisiones Biología y

Computación del Departamento de Ciencias Básicas y doce alumnos de las carreras de: Licenciatura en Sistemas de Información, Licenciatura en Ciencias Biológicas e Ingeniería Agronómica.

Con la incorporación de los estudiantes se pretende lograr su inserción en el campo de la investigación y una completa formación académico-científica. Además, se promoverá la formación de graduados en áreas afines con el fin de incentivar la investigación, principalmente en el área de Sistemas de la Información. De esta manera, podrán involucrarse en Becas de postgrado de la Secretaría de Ciencia y Técnica de la UNLu y otras Instituciones del sistema científico nacional. Esto conllevará a la realización de tesis de grado y postgrado en temáticas interdisciplinarias entre la Botánica y la Informática.

Referencias

- Beltrán, J. (2005) Manejo de información de biodiversidad en la Argentina: estado de situación. I Reunión de la Red CYTED para la Conservación e Informatización de Colecciones Biológicas - IV Reunión de la Red Nacional de Colecciones Biológicas de Argentina.
<http://www.gbifargentina.org.ar/reunion/contenido/RBeltran.php> (consultado 3-3-14). Buenos Aires, Argentina.
- Carreras, M.E.; Ruiz, G.M. & Bossa, S.R. (2008) Plantas nativas con posible valor ornamental. Aplicación multimedia interactiva. 1º Edic. Sima Editora. Córdoba, Argentina. Versión CD.
- Cerana, M. M.; Gil, S. P. ; Pascualides A.L. & Seisdedos, L. (2007) Atlas de Anatomía de plantas silvestres y cultivadas que crecen en

- Córdoba. Priemra parte: Histología. Editora Sima. Córdoba, Argentina. Versión CD.
- Cosa, M. T; Dottori, N.; Stiefkens, L.; Matesevach, M.; Delbón, N. ; Weimer, P.; Machado, S.; Figueroa, S. & Cabrera, V. (2012) Atlas de Anatomía Vegetal. Estructuras secretoras. Cátedra de Morfología Vegetal. Escuela de Biología. Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional del Córdoba. Versión CD.
- Cosa, M. T; Dottori, N.; Hadid, M.; Stiefkens, L.; Matesevach, M.; Delbón, N.; Weimer, P.; Machado, S. & Figueroa, S. (2013) Atlas de Anatomía Vegetal. Adaptaciones de las plantas vasculares. Cátedra de Morfología Vegetal. Escuela de Biología. Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional del Córdoba. Versión CD.
- Dottori, N.; Cosa, M. T; Bruno, G.; Hadid, M.; Stiefkens, L.; Delbón, N & Matesevach (2008) Atlas de Anatomía Vegetal II. Estructuras Reproductivas. Cátedra de Morfología Vegetal. Escuela de Biología. Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional del Córdoba. Versión CD.
- FCNyM (Facultad de Ciencias Naturales y Museo de La Plata) (2013) digitalización de especímenes del Herbario de la División de Plantas Vasculares. http://www.museo.fcnym.unlp.edu.ar/articulo/2013/4/24/herbario_notas (consultado 2-3-14). Buenos Aires, Argentina.
- IBODA (Instituto de Botánica Darwinion) (2014) Iris, informatización de la Red Nacional de Herbarios. Herbarios IBODA, IMBIV, IBONE e INTA. www2.darwin.edu.ar/Herbario/Bases/ (consultado 2-3-14). Buenos Aires, Argentina.
- IMBIV (Instituto Multidisciplinario de Biología Vegetal) (2014) Herbario. <http://imbiv.unc.edu.ar/museobotanicoherbario> (consultado 1-3-14). Córdoba, Argentina.
- MACN (Museo Argentino de Ciencias Naturales) (2014) Proyecto de apoyo a la informatización de las colecciones del MACN http://www.macn.secyt.gov.ar/investigacion/proyectos/colecciones/coleccionesfw/pro_colecciones_pvasculares.php (consultado 1-3-14). Buenos Aires, Argentina.
- Mariño, S. I., Tressens, S. G. & Pace, G. J. (1998) Informatización de datos florísticos de la Reserva Natural Provincial del Iberá (Corrientes - Argentina). Reunión de Comunicaciones científicas y Tecnológicas. Actas Tomo III. Ciencias Agrarias - Ciencias Biológicas. Ctes. Arg. 6-103 a 6-105.
- Mata Montero, E. & Mata, E. (2006) Informatización de conocimiento para la conservación de la biodiversidad. Ecosistemas 15(2): 58-65.
- Ramez, E. & Shamkant, N. (2007) Fundamentos de Sistemas de Bases de Datos. Pearson Education. 5ta. Edición. Reino Unido.
- SNDB (Sistema Nacional de Datos Biológicos) (2014) <http://www.sndb.mincyt.gob.ar/> (consultado 2-3-14).

Estudio, análisis y diseño de estrategias de diseño de software colaborativo

Rodrigo Santos¹, Leonardo Ordinez², Damián Barry², Sergio Firmenich², Nahuel Defossé², Alejandro Sánchez², Roberto Voogt²

1 Dep. de Ing. Eléctrica y Computadoras, Instituto de Inv. en Ingeniería Eléctrica, Universidad Nacional del Sur
Av. Alem 1253 Bahía Blanca, (0291) 4595101
ierms@criba.edu.ar

2 Laboratorio de Investigación en Informática (LINVI), Departamento de Informática, UNPSJB
Bvd Brown 2051 Puerto Madryn, (0280) 4883585
leo.ordinez@gmail.com, damian_barry@unpata.edu.ar, sergio.firmenich@gmail.com, nahuel.defosse@gmail.com, aljsanchez@gmail.com, robertoc.voogt@gmail.com

Resumen

La motivación principal de este proyecto se asienta en el área de aplicación de los sistemas a desarrollar, los cuales tienen que ver con situaciones complejas, multicausales, posiblemente extraordinarias y que involucran la interacción de diferentes actores, en un espacio geográfico extenso, con potencialmente bajo nivel de conectividad. Existen diversos contextos que presentan dichas condiciones, como son la extensión del terreno, la recolección de datos, su sistematización, el almacenamiento y procesamiento correspondientes, entre otras. A pesar de tener una fuerte impronta experimental, para de lograr la extrapolación de los conocimientos adquiridos mediante la experiencia, es necesario conceptualizar el problema abordado, enmarcarlo y modelarlo para poder elaborar conclusiones generales que permitan obtener métodos y técnicas de carácter

genérico al más amplio espectro de situaciones. En este sentido, la particularidad geográfica de la Patagonia es análoga en el contexto marítimo, el bosque nativo andino-patagónico, el ámbito turístico, la generación de energía y al contexto vial, entre otras.

Unas áreas de aplicación destacadas, de los conceptos generales mencionados antes, tienen que ver con el monitoreo comunitario y la ciencia ciudadana. Con el objetivo de lograr la participación ciudadana en ambas alternativas y amenizar la tarea de recolección, se plantea la posibilidad de incluir conceptos de ludificación.

Palabras clave: sistemas colaborativos, ciencia ciudadana, monitoreo comunitario, internet de las cosas.

Contexto

El Linvi es el Laboratorio de Investigación en Informática dependiente

de la Facultad de Ingeniería de la Universidad Nacional de la Patagonia San Juan Bosco. La especialidad del laboratorio es la investigación, desarrollo, servicios de vinculación y formación de recursos humanos en Informática. Las principales líneas de investigación y desarrollo del Laboratorio, tanto en ciencias básicas como aplicadas son las siguientes: Ingeniería de Software, Sistemas de Tiempo Real, Sistemas Distribuidos, Bases de Datos, Inteligencia Artificial y Procesamiento de Imágenes .

El Laboratorio de Sistemas Digitales (LSD), pertenece al Departamento de Ingeniería Eléctrica y de Computadoras de la Universidad Nacional del Sur y al Instituto de Investigaciones en Ingeniería Eléctrica con doble dependencia con la Universidad Nacional del Sur y el CONICET. El grupo está conformado por dos Investigadores Adjuntos del CONICET que además son profesores de las carreras de Ingeniería Electrónica, Electricista y en Sistemas. Todos ellos con título de postgrado. Participan además estudiantes que realizan su Proyecto final de carrera (Tesina) para graduarse de Ingenieros Electrónicos.

Ambos Laboratorios cuentan con Convenios de Colaboración mutua y han realizado trabajos conjuntos en temáticas relacionadas con los sistemas de tiempo real, inteligencia artificial, redes y procesamiento de imágenes. Asimismo, cuentan con un contacto e intercambio permanente de docentes e investigadores que dictan cursos y realizan estancias cortas de investigación en forma recíproca. Por otra parte, cada uno de los Laboratorios en forma individual posee vinculaciones con otros grupos de investigación tanto a nivel nacional como internacional.

Introducción

En el plano de las tecnologías de la información y la comunicación (TIC), un sistema de información efectivo es el que provee a sus usuarios con la información necesaria para responder a los requerimientos medioambientales [El-Bibany, 1991]. La tecnología web en todo el mundo y las redes electrónicas han creado un ambiente, donde el lugar físico no es importante. A partir de las últimas dos décadas, los avances en las TIC han creado capacidades accesibles para integrar y coordinar información de distintos actores y construir conocimiento a partir de ella. En este punto es importante distinguir entre datos, información y conocimiento [Bellinger, 2004]. Los datos son hechos del mundo físico que se obtienen mediante observación. La información agrega el contexto necesario a los datos para poder obtener un sentido de los mismos. Aunque la información puede ser utilizada para tomar decisiones, por sí sola no es suficiente. Se requiere de aprendizajes previos y experiencia para interpretar esa información (datos contextualizados). Dicho bagaje, junto con la información correspondiente, constituye el conocimiento, el cual permite constituir un juicio sobre la realidad y tomar acciones sobre la misma.

La motivación principal de la propuesta se asienta en el área de aplicación de los sistemas a desarrollar, los cuales tienen que ver con situaciones complejas, multicausales, posiblemente extraordinarias y que involucran la interacción de diferentes actores, en un espacio geográfico extenso, con potencialmente bajo nivel de conectividad.

Desde un punto de vista más abarcativo, la implementación de un sistema colaborativo para la generación de conocimiento, permitirá la obtención de un grado de comprensión superior al

del conocimiento: la sabiduría [Bellinger, 2004]. Este nivel de entendimiento posibilita la reflexión sobre el conocimiento obtenido, lo cual deriva en nuevos conocimientos a un nivel más general. En particular, se pretende obtener experiencia, capaz de ser generalizada y aplicada, para la construcción de otros sistemas de similares características que aborden otros contextos. No obstante, dichos contextos conserven ciertas condiciones análogas a las que se encuentran en este proyecto, como son la extensión del terreno, la recolección de datos, su sistematización, el almacenamiento y procesamiento correspondientes, entre otras. A fin de lograr esta extrapolación de los conocimientos adquiridos mediante la experiencia, es necesario conceptualizar el problema abordado, enmarcarlo y modelarlo para poder elaborar conclusiones generales que permitan obtener métodos y técnicas de carácter genérico al más amplio espectro de situaciones. En este sentido, la particularidad geográfica de la Provincia del Chubut es análoga en el contexto marítimo, el bosque nativo andino-patagónico, el ámbito turístico en general, la generación de energía en diferentes formas (hidráulica, térmica, eólica y solar) y al contexto vial, entre otras.

Líneas de Investigación, Desarrollo e Innovación

El proyecto se inicia en el año 2016 y se plantean los siguientes ejes de actividades.

Actividades de Investigación y desarrollo

1. Seleccionar material bibliográfico y generar una base de conocimiento sobre

las técnicas y métodos empleados en Sistemas Colaborativos y afines (Computación Móvil, Internet de las Cosas, Redes Oportunistas, Ciencia Ciudadana, etc.).

2. Investigación de tecnologías adecuadas para el desarrollo de Sistemas Colaborativos, Computación Móvil, Internet de las Cosas, Ciencia Ciudadana y otros relacionados.

3. Definir métricas que permitan obtener conclusiones relevantes respecto a las técnicas y métodos implementados.

4. Definir y desarrollar uno o varios métodos de pruebas de clasificación supervisada y no supervisada sistematizadas para someter a comparación las distintos métodos propuestos.

5. Armar un banco de pruebas que permita comprobar las distintas implementaciones y métodos utilizados para el desarrollo de Sistemas Colaborativos. Según la observación de los resultados obtenidos, proponer mejoras y/o reformulaciones.

Actividades Académicas

1. Publicar en congresos, conferencias, revistas y libros de orden nacional e internacional, los resultados obtenidos.

2. Tomar contacto con grupos de investigación en la disciplina, de origen nacional e internacional, participando en proyectos en conjunto o mediante el intercambio de integrantes de los proyectos.

3. Realizar actividades de posgrado: cursos, seminarios, conferencias, reuniones, charlas técnicas y talleres.

4. Incrementar el número de participantes tanto de docentes, alumnos y graduados en los proyectos de investigación de la sede.

5. Difundir y transferir al medio. Insertarse en la sociedad como un grupo

de investigación, desarrollo e innovación (I+D+i), creador de recursos humanos altamente calificados como actores en el desarrollo tecnológica de la región.

Resultados y Objetivos

Objetivo General

Estudiar, analizar y desarrollar estrategias y técnicas de diseño de sistemas colaborativos para la generación de conocimiento acerca del ambiente.

Objetivos Específicos de Investigación y Desarrollo

- A. Investigar y comparar metodologías de desarrollo de software general, adaptables a sistemas colaborativos.
- B. Analizar tecnologías apropiadas para la construcción de sistemas colaborativos en diferentes contextos ambientales.
- C. Proponer estrategias originales o innovaciones a las existentes para la gestión de proyectos de software en las áreas afines a los sistemas colaborativos: Computación Móvil, Internet de las Cosas, Ciencia Ciudadana, entre otras.
- D. Validar técnicas y estrategias para el desarrollo de aplicaciones de software concretas, en base a tecnologías apropiadas, en diferentes contextos.

Objetivos Específicos Académicos

- A. Constituir mediante el proyecto, un grupo de investigación de la Universidad Nacional de la Patagonia San Juan Bosco sede Puerto Madryn, sobre la disciplina Sistemas Colaborativos y afines (Computación Móvil, Internet de las Cosas, Redes Oportunistas, etc.).

- B. Fomentar, incentivar y difundir las tareas de investigación.
- C. Mejorar la formación de recursos humanos altamente calificados, con capacidades de investigación y desarrollo. Lograr la categorización de los docentes participantes y la jerarquización del departamento de informática y de la universidad en todos sus niveles.
- D. Consolidar las actividades realizadas por el Laboratorio de Investigación en Informática (LINVI) perteneciente al Departamento de Informática de la UNPSJB.
- E. Interactuar con otros grupos de investigación de las sedes de la universidad y de otras universidades, en tareas conjuntas de investigación y desarrollo, como también en la formación de recursos humanos.
- F. Incrementar el número de proyectos acreditados y de trabajos publicados por la universidad y la sede.

Formación de Recursos Humanos

Por ser un proyecto de investigación nuevo en la Sede, aún no se cuentan con tesis o tesinas en curso o finalizadas. No obstante, se estima que a lo largo de este año se iniciarán algunas.

Referencias

1. Bellinger, Gene; Durval Castro; Anthony Mills (2004). "Data, Information, Knowledge, and Wisdom", <http://www.systems-thinking.org/dikw/dikw.htm>.
2. Bava, J.O., Rechene, C., 2004. Dinámica de la regeneración de lenga (*Nothofagus pumilio* (Poepp. et Endl.) Krasser) como base para la aplicación de sistemas silvícolas, in: Arturo, M., Franji,

- J., Goya, J. (Eds.), *Ecología Y Manejo de Los Bosques*. EDULP, La Plata.
3. Defossé, G.E., Godoy, M.M., Bianchi, L.O., Lederer, N.S., Kunst, C., 2015. Fire History, Fire Ecology and Management in Argentine Patagonia: From Ancient Times to Nowadays, in: Leblon, B., Alexander, M.E. (Eds.), *Current International Perspectives on Wildland Fires, Mankind and the Environment*. Nova Science Publishers, pp. 177–210.
 4. Ferretti, M., 2013. Forest Monitoring: an introduction, in: *Forest Monitoring - Methods for Terrestrial Investigations in Europe with an Overview of North America and Asia, Developments in Environmental Science*. Elsevier, pp. 3–18.
 5. López Bernal, P.M., Defossé, G.E., Quinteros, P.C., Bava, J.O., 2012. Sustainable Management of Lenga (*Nothofagus pumilio*) Forests Through Group Selection System, in: Garcia, J.M., Diez Casero, J.J. (Eds.), *Sustainable Forest Management - Current Research*. INTECH, pp. 45–66.
 6. SAyDS, 2005. *Primer inventario nacional de bosques nativos*.
 7. Veblen, T.T., Kitzberger, T., Raffaele, E., Mermoz, M., González, M.E., Sibold, J.S., Holz, A., 2008. The historical range of variability of fires in the AndeanPatagonian *Nothofagus* forest region. *Int. J. Wildl. Fire* 17, 724–741.
 8. "Emerging technologies: Smarter ways to fight wildlife crime" - Emerging technologies / Environmental Development 12 (2014) 62 – 72
 9. Rajiv Ranjan, Samee Khan, Joanna Kolodziej, Albert Y. Zomaya "Guest Editors' Introduction: Cloud-Based Smart Evacuation Systems for Emergency Management", *IEEE Cloud Computing*, Volume 1, Number 4, November 2014
 10. Karan Mitra, Saguna, Christer Åhlund, "A Mobile Cloud Computing System for Emergency Management", *IEEE Cloud Computing*, Volume 1, Number 4, November 2014
 11. Mianxiong Dong, He Li, Kaoru Ota, Laurence T. Yang, Haojin Zhu, "Multicloud-Based Evacuation Services for Emergency Management", *IEEE Cloud Computing*, Volume 1, Number 4, November 2014
 12. SERGIO F. OCHOA; RODRIGO SANTOS, *Human-centric Wireless Sensor Networks to Improve Information Availability During Urban Search and Rescue Activities*, *INFORMATION FUSION*; Lugar: Amsterdam; Año: 2015 vol. 22 p. 71 - 84.
 13. JAVIER OROZCO; RODRIGO SANTOS; SERGIO F. OCHOA; LEONARDO ORDINEZ; ROC MESEGUER; NELSÓN BALOIAN, *A clinically assisted collaborative system architecture for preventing falls in elderly people* *LECTURE NOTES IN COMPUTER SCIENCE*; Año: 2014
 14. ALVARO MONARES; SERGIO F. OCHOA; RODRIGO SANTOS; JAVIER OROZCO; ROC MESEGUER *Modeling IoT-Based Solutions Using Human-Centric Wireless Sensor Networks* *SENSORS*; Lugar: Basel; Año: 2014 vol. 14 p. 15687 - 15713.

Implementación de un sistema de video conferencia HD autónomo utilizando Raspberry Pi

Daniel A Giulianelli, Rocío A Rodríguez, Pablo M Vera, Anabella G. Conca,
Gabriela Y Vallés, Pablo Cammarano

GIDFIS (Grupo de Investigación, Desarrollo y Formación en Innovación de Software)
Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza
Florencio Varela 1903, San Justo, Provincia de Buenos Aires
gidfis@ing.unlam.edu.ar

Resumen

Esta línea de I+D (Investigación y Desarrollo), está enfocada en la construcción de un software que permita realizar video-conferencias tanto sobre redes IPv4 como IPv6. Para este objetivo se ha considerado una solución autónoma que pueda ser conectada a un monitor o televisor, para lo cual también se proveerá la parte de hardware mediante el uso de Raspberry PI a las que se conectarán módulos de cámara a través del puerto SCI las cuales tienen mejor velocidad que las cámaras USB y permitirán tener una resolución HD. En uno de los nodos, se trabajará además en la motorización de la cámara de forma que a través de sensores de sonido pueda automáticamente seguirse al orador. Las principales ventajas de esta propuesta son el bajo costo frente a las soluciones existentes en el mercado, la compatibilidad con redes IPv6 y la autonomía de la solución en donde el usuario final no tiene que instalar un software ni tener una computadora la cual tenga ciertas prestaciones.

Palabras clave: Video-Conferencia, IPv6, Raspberry Pi, Cámara Motorizada

Contexto

El presente proyecto de Investigación y Desarrollo corresponde al programa PROINCE. Siendo llevado a cabo por el GIDFIS (Grupo de Investigación, Desarrollo y Formación en Innovación de Software) perteneciente al Departamento de Ingeniería e Investigaciones Tecnológicas. El grupo es interdisciplinar está formado por docentes de distintos departamentos y en él se encuentran formándose en actividades de I+D alumnos de la universidad.

Introducción

En esta línea de investigación se plantea el desarrollo de un software que está diseñado para un hardware específico. Los SBC (Single Board Computer) “son computadores completos en una sola placa de tamaño reducido, que son lo suficientemente potentes para ejecutar un sistema operativo real, además estos son de bajo costo y versátiles. Por lo cual son dispositivos útiles para la ejecución de diversos proyectos ya sean educativos o de investigación científica”[1]. Una Raspberry Pi 2 Model B [2] tiene las características en cuanto a hardware que se presentan en la tabla 1.

Tabla 1. Características Raspberry PI 2 Model B

RAM	1 GB
Almacenamiento	Micro SD
Procesador	ARM A7
Velocidad	900 Mhz
Ethernet	Sí
Salida HDMI (audio y video)	Sí
Nº puertos USB	4
Soporta instalación de S.O.	Sí (Linux)
Cantidad de GPIO (puertos útiles)	40
Lenguajes de programación soportados	Python, C, C++, Java, Perl, Ruby, etc
Alimentación	5v
Tamaño	85 x 56 x 17mm

En una memoria SD es posible instalar un sistema operativo como Linux y luego un programa que permita realizar algo en particular, en el caso del presente proyecto una aplicación de video-conferencia, desarrollada específicamente para este hardware.

El sistema de video conferencia permitirá ser utilizado en IPv6 aprovechando sus ventajas, pero también permitirá su uso en IPv4.

A nivel académico IPv6 es un tema en

constante desarrollo. Algunos artículos han tratado la temática de transición de IPv4 a IPv6 mediante el estudio de métodos de acceso transparente como es el caso del trabajo [3]. Se continúa trabajando en la implementación y despliegue de IPv6 en redes académicas [4], [5]. También el estudio de características puntuales como por ejemplo: multicast, calidad de servicio resulta de interés para diversos autores entre ellos [6], [7]; así como cuestiones de movilidad sobre IPv6 [8], [9], [10].

A nivel mundial es notable la adopción de IPv6 y su crecimiento marcado, ver figura 1 (extraída de [11]). Un 8,33% de los accesos se realizan por IPv6 siendo actualmente, en su totalidad, en forma nativa.

A nivel mundial [12] los 10 países con mayor porcentaje de adopción de IPv6 son: Bélgica (40,01%), Estados Unidos (23,49%), Portugal (22,78%), Alemania (21,65%), Grecia (21,22%), Perú (16,23%), Ecuador (13,44%), Estonia (10,12%), Japón (9,87%), Malasia (8,93%). No obstante en Argentina el porcentaje de adopción de IPv6 es bajo, representando tan solo un 0,03%.

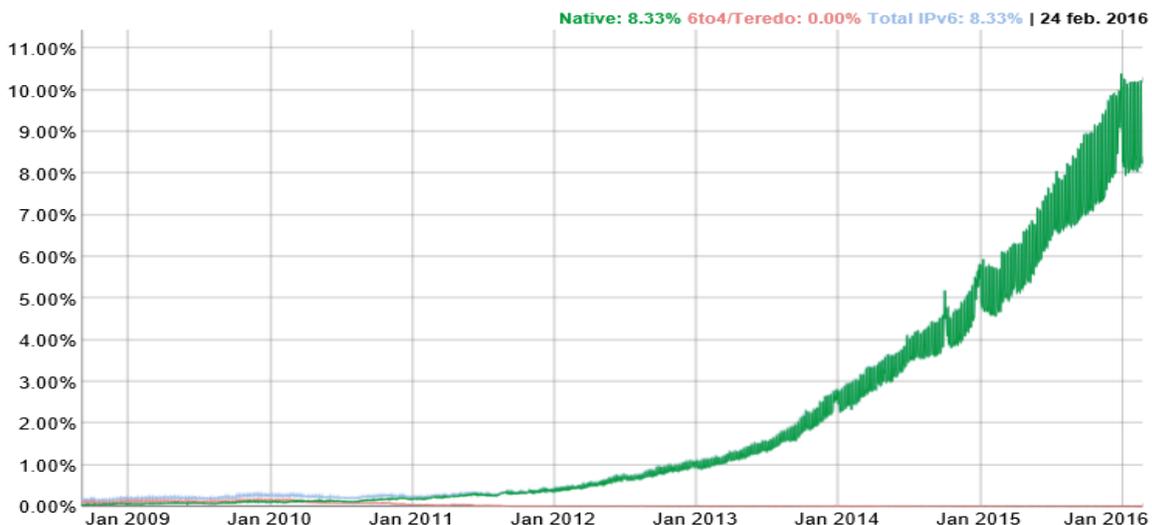


Figura 1. Crecimiento de Accesos Mediante IPv6

En la tabla 2 se muestra el grado de adopción los países de América. Como era de esperarse en base a la referencia mundial Estados Unidos, Perú y Ecuador son los que tienen los mejores porcentajes en el continente americano, seguidos por Brasil (6,82%), Canadá (6,77%), Bolivia (3,29%), Trinidad y Tobago (1,03%). Siendo estos 7 países los que destacan del resto quienes no alcanzan el 1%. Como puede observarse en la tabla 3, hay 8 países sin adopción de IPv6.

Tabla 2. Porcentaje de Adopción de Ipv6 en América

	País	Porcentaje
1	Estados Unidos	23,49
2	Perú	16,23
3	Ecuador	13,44
4	Brasil	6,82
5	Canadá	6,77
6	Bolivia	3,29
7	Trinidad y Tobago	1,03
8	Belice	0,86
9	Puerto Rico	0,54
10	Guyana	0,07
11	Guatemala	0,06
	México	0,06
12	Honduras	0,05
13	Argentina	0,03
	Colombia	0,03
14	Chile	0,02
	Nicaragua	0,02
	Uruguay	0,02
15	Costa Rica	0,01
	Haití	0,01
	Paraguay	0,01
	República Dominicana	0,01
	Surinam	0,01
	Venezuela	0,01
16	Bahamas	0,00
	Cuba	0,00
	El Salvador	0,00
	Groenlandia	0,00
	Guayana Francesa	0,00
	Jamaica	0,00
	Martinica	0,00
	Panamá	0,00

La tabla 2 está ordenada por porcentaje lo cual permite generar un ranking, en el cual pueden verse los 7 primeros puestos conformados por los países mencionados previamente, el puesto 11, 13, 14, 15 son compartidos por dos ó más países; quedando en el puesto 16 aquellos países que no tienen adopción a IPv6 siendo el porcentaje 0%. Estando Argentina junto a Colombia en el puesto 13.

Por este motivo, si bien la aplicación a desarrollar tendrá soporte para IPv6 también funcionará en IPv4 dado que el porcentaje de adopción a IPv6 aún es bajo.

Líneas de Investigación, Desarrollo e Innovación

Los ejes principales de trabajo son:

- Análisis de Técnicas de Streaming de Video sobre IPv6
- Implementación de Software que incorpore Streaming de Video sobre IPv6
- Manejo de Hardware dedicado (sensores de sonido y ultrasonido, servomotores)
- Implementación con Raspberry Pi

Resultados y Objetivos

Actualmente se ha avanzado en la adquisición de todo el hardware necesario para armar el prototipo, planificado inicialmente para poder conectar 4 nodos los cuales tendrán el mismo hardware y aplicación instalada. Cada uno de estos nodos está implementado con una solución de hardware que incluye una Raspberry Pi 2 Model B.

Cada una de ellas se conectará por HDMI a un monitor o televisor, se le agregó a la solución una cámara. Existen dos opciones de cámara para conectar a la Raspberry Pi: (A) a través del puerto USB

ó (B) usando el puerto SCI de la Raspberry Pi (ver figura 2).

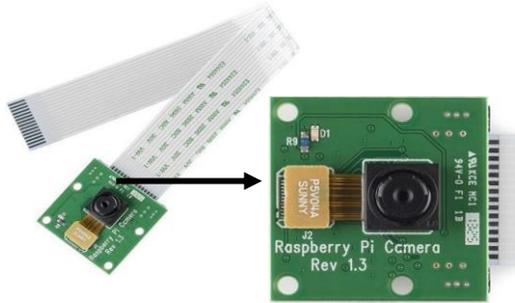


Figura 2. Módulo de Cámara – Puerto SCI

En el caso de la cámara USB, la ventaja que ofrece es la de poder programar una aplicación desde la PC de escritorio utilizando la cámara, y luego conectar esa cámara y llevar la aplicación a la Raspberry sin tener que realizar modificaciones. Mientras que la ventaja de la cámara integrada mediante el puerto SCI es la velocidad y la calidad de captura de imagen y video. El gabinete de la Raspberry trae un zócalo en donde se inserta la placa correspondiente al módulo de la cámara (ver figura 3).



Figura 3. Módulo de Cámara conectado a la Raspberry Pi 2 Model B

Se incorporan diversos módulos y elementos adicionales a la Raspberry Pi:

- Módulo de WIFI por el cual se establecerá la conexión a internet.
- Sensores (sonido y ultrasonido), los cuales se conectan a un puerto denominado GPIO (General Purpose input/output) especialmente diseñado para operaciones de entrada/salida de propósito general. Este puerto tiene distintos

pins que pueden utilizarse para controlar fácilmente tanto sensores como actuadores (ver figura 4).

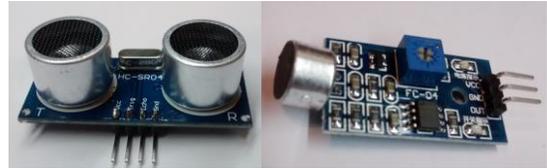


Figura 4. Sensor sonido y ultrasonido

- Motorización, los sensores permitirán detectar la posición del orador dentro de una sala y motorizar la cámara para que pueda moverse siguiendo al orador. Para mover la cámara se agregan dos servomotores (vertical y horizontal) los cuales moverán la cámara (ver figura 5).

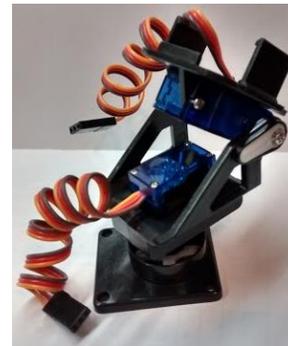


Figura 5. Servomotores instalados

Actualmente se está trabajando en mayor medida con la parte del software que dará soporte a la video-conferencia en el prototipo diseñado. Tomando en cuenta no sólo el desarrollo de la aplicación que cuente con soporte a IPv6 e IPv4 sino también analizar la forma en que el usuario podrá interactuar con la aplicación, seteando inicialmente la red wifi a utilizar, eligiendo los usuarios con los que conectarse, etc. Para lo cual se planificó utilizar un teclado que tiene mouse incorporado (diseñado para los Smart TV). Ha sido probado en las Raspberry y funciona correctamente, entonces las

interfaces deben ser planificadas para que sean sencillas de usar con estos teclados ya adquiridos.

Formación de Recursos Humanos

El equipo de trabajo está compuesto por 14 personas, 9 docentes y 5 alumnos de las carreras de Ingeniería Electrónica e Informática; los cuales en su mayor parte cuentan con becas asociadas a sus actividades de I+D. En relación con esta línea de investigación se encuentran asociadas: 2 tesinas de grado y 1 tesis de maestría.

Referencias

- [1] Gonzalez, K., D. Urrego, W. Gordillo. (2016). "Estudio sobre Computadores de Placa Reducida Raspberry Pi Modelo B y Cubieboard2 en la Creación de Proyectos con Base Tecnológica de Gran Impacto Social." ENGI Revista Electrónica de la Facultad de Ingeniería 3.1
http://revistas_electronicas.unicundi.edu.co/index.php/Revistas_electronicas/article/download/164/159
- [2] Raspberry. (2016) "Raspberry Pi 2 Model B"
<https://www.raspberrypi.org/products/raspberry-pi-2-model-b/>
- [3] Mercado, Gustavo, et al. (2011). "Implementación y Evaluación de métodos de Traslación de Protocolos para la transición IPv4-IPv6."
<http://hdl.handle.net/10915/18769>
- [4] Cicileo, Guillermo, Mariela Rocha. (2014). "Evolución de la red de interconexión universitaria de Argentina."
<http://hdl.handle.net/10786/778>
- [5] Mercado, Gustavo, et al. (2014). "ACyT-Net Red Académica Científica y Tecnológica Experimental de Mendoza Argentina."
<http://hdl.handle.net/10915/41702>
- [6] Facchini, Higinio Alberto, et al. (2013). "Análisis de prestaciones de tráfico multicast en redes mixtas IPv4 e IPv6."
<http://hdl.handle.net/10915/27085>
- [7] Díaz, Javier F., et al. (2010) "Modelos de QoS en redes IPv6, integración con otras redes."
<http://hdl.handle.net/10915/19420>
- [8] Cetin, G.; Bilisim Enstitusu, Elektron. Ve Bilgisayar Egitimi, Gazi (2012). "Development of an mobile agent platform for IPv6 network environment: Mobile-C", Universidad de Ankara, Turkía.
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6204507
- [9] Taffernaberry, Juan Carlos, et al. (2014). "PMIP6: Análisis, Evaluación y Comparación de ambientes Proxy Mobile IP en versión 6, aplicado a Redes de Avanzada".
<http://hdl.handle.net/10915/41658>
- [10] Taffernaberry C, Tobar S, Mercado G, Noguera J, Perez Montes. (2015) "Aplicaciones del RFC 5213 – Proxy Mobile IPv6 – Comparación de ambientes móviles en redes académicas".
<http://44jaiio.sadio.org.ar/sites/default/files/ieftday17-21.pdf>
- [11] Google. (2016) "Google IPv6. Statistics".
<https://www.google.com/intl/en/ipv6/statistics.html>
- [12] Google. (2016) "Google IPv6. Per-Country IPv6 Adoption"
<https://www.google.com/intl/en/ipv6/statistics.html#tab=per-country-ipv6-adoption&tab=per-country-ipv6-adoption>

Gestión, Inventario y Monitoreo Hardware con Alertas Automáticas (G.I.M.H.A.A.)

Silvia Edith Arias, Fabián Gibellini, Analía L. Ruhl, M. Alejandra Di
Gionantonio, Nora V. Flores, Mónica M. Serna, Daniel F. Arch,
Ezequiel Ambrogio, Milagros Zea Cárdenas, Germán Parisi,
Diego Barrionuevo.

Laboratorio de Sistemas / Dpto. de Ingeniería en Sistemas / Universidad
Tecnológica Nacional / Facultad Regional Córdoba

Cruz Roja S/N, 5016

s_autn@hotmail.com, fgibellini, lruhl@bbs.frc.utn.edu.ar, ing.alejandradg, ingnoraflores,
sernamonicam@gmail.com, daniel.arch@pjn.gov.ar, ezequielambrogio, milyzc, germannparisi,
santosdiegob@gmail.com

Resumen

El Laboratorio de Ingeniería en Sistemas de Información (LabSis) de la Universidad Tecnológica Nacional - Facultad Regional Córdoba (U.T.N. - F.R.C.) dispone de aulas con un amplio parque informático para distintas actividades correspondientes a carreras de grado, cursos o diplomaturas. Debido al incremento que cada año este parque tiene, surgió la necesidad de mejorar el control en las PCs instaladas en cada de sus aula y que son de uso frecuente tanto por docentes como por el alumnado de la U.T.N. - F.R.C.. En este marco, está en desarrollo un sistema de información que permite monitorear las PCs dentro de las subredes existentes, generando alertas de forma automática ante comportamientos indeseados, reemplazos o sustracción de los mismos e informando a los usuarios responsables, mediante EMAILS, SMS, la WEB, o alguna otra vía configurable. Solo se notifican los eventos estipulados como no acordes a la correcta distribución, organización y funcionamiento de estos recursos tecnológicos. Este sistema será software

libre, ampliamente escalable por lo que podrá ser implementado dentro de cualquier red informática.

Palabras clave: Alertas, Inventario, Hardware, Monitoreo, Historial, Gestión

Contexto

Últimamente las organizaciones han incrementado la cantidad de sus activos informáticos, y esta es una tendencia que sigue en aumento ya sea debido a la necesidad de procesamiento, nuevos puestos de trabajo o porque actualmente la mayoría de las organizaciones requieren que cada uno de sus trabajadores desempeñe sus funciones en una estación de trabajo, ya sea que esta posea almacenamiento y/o procesamiento o que trabaje en la nube. En el ámbito académico esta tendencia también es válida, ya sea por las causas antes mencionadas o porque el número de ingresantes aumenta con los años lo que lleva a incrementar el parque tecnológico de los laboratorios o gabinetes informáticos, en particular las instituciones que dictan carreras afines a la computación.

Este incremento o proliferación de activos de TI acentúa la complejidad del entorno [1] de TI en relación a la Gestión de Incidencias, Gestión de Problemas (cuando un incidente se vuelve recurrente) y Gestión de la Configuración y Activos TI del Servicio [2]. Lo que lleva a la mayoría de las organizaciones a atender los incidentes cuando justamente ya han pasado a ser un incidente o peor aún, un problema debido a que no se puede gestionar correctamente lo que se desconoce y en muchos casos un activo TI inicia a un proceso de mantenimiento porque se ha manifestado un incidente con el mismo.

Las recomendaciones de gestión de activos ITIL [3] propone la actividad de Gestión de Configuración y Activos TI del Servicio que lleva un registro actualizado de todos los elementos de su infraestructura TI junto con sus interrelaciones.

Dentro de elementos podemos mencionar [2]:

- Dispositivos de hardware como PCs, impresoras, routers, monitores, etc. así como sus componentes: tarjetas de red, teclados, etc.
- Software: sistemas operativos, aplicaciones, protocolos de red, etc.
- Documentación: manuales, acuerdos de niveles de servicio, entre otros.

El Laboratorio de Sistemas (LabSis) de la Universidad Tecnológica Nacional - Facultad Regional Córdoba (U.T.N. - F.R.C.) tiene como principal objetivo ofrecer disponibilidad y un correcto funcionamiento de los equipos (PCs) a alumnos y a docentes para lo que implementará un sistema de inventario de los componentes de los equipos con carga automatizada, además se monitoreará del estado de cada componente junto con un módulo de alertas automáticas que se generarán al identificar algún cambio de

estado no deseado en algún componente. Alineado con el objetivo de LabSis, este sistema contemplará todos las PCs dentro de cada subred que conforman a cada una de las aulas habilitadas en LabSis. Esto permitirá dar soporte a una Gestión de Configuración y Activos TI del Servicio dentro de estas subredes teniendo mayor control, gestión y conocimiento sobre los equipos que se ponen a disposición y lograr una mejoría continua de los servicios brindados a alumnos y docentes.

El presente trabajo está inserto dentro del proyecto homologado por la Secretaría de Ciencia y Técnica de la Universidad Tecnológica Nacional “Inventario Seguro en Ambiente Informáticos con Alertas Automáticas y Registro de historial para activos de TI” - Código: EIUTNCO0002226, radicado en el Laboratorio de Ingeniería en Sistemas de Información (LabSis) de la U.T.N. - F.R.C., quien es a su vez es el organismo que contribuye a su financiamiento.

Introducción

LabSis surge poco después de haber comenzado las carreras de informática en la U.T.N – F.R.C. Inicialmente contaba con escasos recursos que, gradualmente, se fueron incrementando hasta poseer una gran plataforma tecnológica, compuesta de estaciones de trabajo, servidores y dispositivos que permiten la administración de su red interna.

La principal función de LabSis es ofrecer disponibilidad de las estaciones de trabajo a alumnos y a docentes para:

- El dictado cotidiano de distintas cátedras de la carrera de grado Ingeniería en Sistemas de Información, Diplomaturas o cursos dictados por la U.T.N. - F.R.C.
- La toma de exámenes finales de diferentes cátedras, que implica un proceso de respaldo de los datos para

su posterior corrección por los docentes.

- Competencias de Programación.
- Olimpiada Informática Córdoba (O.I.C.) [4].
- Prácticas libres, llamadas así porque se efectúan durante el tiempo libre de los alumnos.

Este laboratorio ha tenido un crecimiento constante de su plataforma tecnológica, esto se debe a que la cantidad de alumnos ingresantes se ha incrementado, como así también los distintos convenios que la U.T.N. - F.R.C. logra con otras entidades para ofrecer cursos, capacitaciones, desarrollos de software, etc. Para cumplir con su objetivo, LabSis utiliza sistemas automatizados que permiten la optimización de sus procesos y protocolos llevados a cabo en cada situación.

Es por ello que, en consecuencia a lo anteriormente mencionado, la gestión de incidentes (comportamientos no deseados, reemplazos o sustracciones) se ha modificado hasta adoptar el procedimiento actual, en la que se notifica a integrantes del Área Técnica cuando una estación de trabajo o servidor manifiesta un incidente que impida su normal desempeño. Ante esta situación esta área interviene, realizando pruebas sobre el equipo para detectar los componentes afectados, quedando el equipo fuera de cualquier uso hasta la adquisición o reparación de los componentes dañados. Este procedimiento es adoptado por varios laboratorios, como también por empresas y organizaciones.

Lo planteado anteriormente expone la necesidad de conocer en cualquier momento: ¿Qué PCs están habilitadas? ¿Cuáles de sus componentes funcionan correctamente y cuáles no? para saber con anterioridad a sus posibles usos, si surgió o no, alguna incidencia en las mismas.

Por ejemplo cuando la placa de red deja de funcionar existe un cambio de estado en este componente. Estos cambios pueden representar en el inicio de un proceso de reparación para el equipo (correctivo o de mantenimiento) y al mismo tiempo este proceso puede provocar más cambios en un equipo, por ejemplo el reemplazo de una placa de memoria por una nueva.

Dentro de este ambiente académico siempre está la necesidad de lograr una mejora continua de los servicios brindados. En este caso se busca dar soporte a una Gestión de Configuración, recomendadas por ITIL y enfocada hacia dispositivos de hardware, específicamente a las PCs (y sus componentes) que están distribuidas dentro de distintas subredes que conforman cada una de las aulas de LabSis para generar información valiosa de las mismas. Esta información permitirá un mejor control, gestión y conocimiento del estado de las PCs en cierto momento, siendo este activo el recurso material más usado del laboratorio. La información generada alimentará la planificación de una mantención preventiva con la que se buscará una constante optimización de la productividad de los activos TI y a largo plazo cierta predictibilidad en el comportamiento de este recurso.

Para mejorar la toma de decisiones respecto a este hardware se identificaron las siguientes metas a alcanzar:

- Monitorear los componentes de las PCs y registrar los cambios identificados en los mismos.
- Llevar un inventario informatizado con carga automática, catalogado y actualizado de ciertos componentes de cada PC en cada aula de LabSis.
- Generar alertas automáticas que notifiquen a los usuarios los cambios ocurridos.

- Registrar todos los cambios ocurridos ya que cada uno representa un posible incidente.
- Lograr una trazabilidad del estado de los activos monitoreados, para alcanzar eventualmente cierta predictibilidad en su comportamiento.
- Dar soporte a la toma de decisiones relacionadas al hardware, como por ejemplo el tiempo promedio de vida útil de cierto dispositivo que permitirá establecer períodos de compra de los mismos, basados en datos reales.
- Lograr una mejor distribución de los costos de inversión en esta área.

Se estudiaron varias herramientas, libres y con licencia paga, que realizaban diferentes inventarios automatizados pero sin alertas automáticas y en algunos casos sin resguardo de un historial de cambios. De todas las estudiadas, la que mejor se adaptaba era el OCS-Inventory [5], el cual en posteriores pruebas realizadas en equipos del LabSis mostraba inconsistencias en los datos sin ningún patrón o criterio reconocible, por lo que terminó siendo descartada.

Líneas de Investigación, Desarrollo e Innovación

La línea de investigación en sistemas de información se centra en el conocimiento en Ingeniería del Software, es decir, en la aplicación de un enfoque o proceso sistemático, disciplinado y cuantificable al desarrollo de software [6], que permita la obtención de resultados óptimos de la forma más eficiente. Logrando un producto innovador, ya sea que se está creando un nuevo producto o se está mejorando alguno existente.

Este proyecto se inscribe dentro de esta línea de investigación, enfocado en la gestión de activos de T.I. y que afecta a

los ámbitos académico, gubernamental y empresarial.

La concreción de este proyecto constituye una posibilidad que favorece ampliamente el crecimiento de la gestión de activos de T.I. principalmente en las instituciones públicas del país, donde se sufren permanentemente sustracciones indetectables.

Al tratarse de un software con una licencia libre [7], todo aquel que desee implementar el sistema podrá acceder a la aplicación y su código, como así también adaptarlo para la estructura del ambiente informático sobre el cual lo desee trabajar.

Resultados y Objetivos

El objetivo del proyecto es desarrollar un sistema integral, web, que sea de libre uso y de fácil acceso para el Laboratorio de Sistemas de la U.T.N. - F.R.C. y para cualquier organismo que lo requiera, haciendo hincapié en brindar un sistema seguro y de muy bajo costo para dar soporte a una Gestión de Configuración y Activos TI del Servicio dentro de estas subredes permitiendo controlar, prevenir, proteger, notificar y tomar decisiones en tiempo y forma sobre los equipos con los que cuenta el laboratorio y lograr una mejora continua de los servicios brindados a alumnos y docentes. Este sistema contemplará todas las PCs dentro de cada subred que conforman a cada una de las aulas habilitadas en LabSis. Para lograr este objetivo el sistema logrará:

- Realizar un inventario de los componentes de cada PC, con carga automatizada.
- Monitorear del estado de cada componente junto con un módulo de alertas automáticas, las cuales se generarán al identificar algún cambio de estado no deseado en algún componente.

- Mantener la trazabilidad de activos de TI, con el reporte diario actualizado del contenido de cada computadora y sus respectivos movimientos, enviando un mail al administrador o a la persona encargada con el detalle. Como así también después de cortes de energía eléctrica, o eventos similares inesperados.
- Brindar información actualizada de cada computadora presente en las aulas, manteniendo una base de datos actualizada con el registro de cambios y/o novedades que detecte esta aplicación.
- Generar informes estadísticos sobre cambios de hardware ocurridos en los equipos, y reportes con información relevante para la toma de decisiones en la Institución a partir de los datos históricos que almacenará el sistema.

El módulo que realiza el monitoreo del estado de cada componente y el módulo de alertas web ya están desarrolladas. Actualmente se está trabajando en el módulo que permitirá tener el inventario con carga de datos automatizada para luego pasar a realizar distintas pruebas hasta que sean posibles la pruebas de aceptación para verificar y validar que el sistema esté en condiciones de ser implementado en su entorno real y definitivo.

Formación de Recursos Humanos

El equipo está compuesto por Profesores investigadores y de apoyo a la investigación, profesores aspirantes a incorporarse a la carrera de investigador y estudiantes investigadores.

Esta línea de trabajo está inmersa en el proyecto homologado y subsidiado por la Secretaria de Ciencia y Tecnología de la U.T.N. – F.R.C. y dado que el objeto de estudio toca transversalmente una gran cantidad de temas de distintas disciplinas

y que los integrantes poseen título de grado en Ingeniería en las diferentes especialidades de Sistemas y de Electrónica; los participantes en este proyecto se formarán y capacitarán en forma flexible y sobre equipos actuales en áreas de interés prioritario tanto a nivel nacional como en el contexto mundial; permitiendo egresados más capacitados que brinden mejores servicios a la industria y a la sociedad.

Se busca también la consolidación del grupo de investigación en el objeto de estudio, mediante la posibilidad de colaborar con el crecimiento profesional de los mismos, los cuales a su vez se ven altamente predispuestos a perfeccionarse en forma continua.

Referencias

[1] “La optimización de TI es una fuente de ventajas competitivas sostenibles”, IBM, 2008, http://www-05.ibm.com/services/es/cio/pdf/CIO_Series_0203_whitepaper.pdf

[2] ITIL: Gestión de Servicios TI, http://itil.osiatis.es/Curso_ITIL/Gestion_Servicios_TI/gestion_de_configuraciones/vision_general_gestion_de_configuraciones/vision_general_gestion_de_configuraciones.php

[3] Recomendaciones ITIL para la Gestión de activos, <http://itilv3.osiatis.es/itil.php>

[4] "Colaboración entre el Gobierno de la Provincia de Córdoba y la UTN - FRC para el desarrollo de Olimpiadas Informáticas", Marciszack, M., Muñoz, R., Castillo, J., Delgado A., Serrano, D. <http://conaiisi.frc.utn.edu.ar/PDFsParaPublicar/1/schedConfs/4/234-650-1-DR.pdf>.

[5] OCS Inventory, <http://www.ocsinventory-ng.org/en/>

[6] “IEEE Standard Glossary of Software Engineering Terminology,” IEEE std 610.12-1990, 1990. ISBN 155937067X.

[7] Free Software Foundation, Página Oficial, <http://www.fsf.org/>

Una Arquitectura Cliente-Servidor para Modelado Conceptual Asistido por Razonamiento Automático

Christian Gimenez¹ Germán Braun^{1,2,3} Laura Cecchi¹
Pablo Fillottrani^{2,4}

email: {christian.gimenez,german.braun,lcecchi}@fi.uncoma.edu.ar, prf@cs.uns.edu.ar

¹*Grupo de Investigación en Lenguajes e Inteligencia Artificial*
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

²*Laboratorio de I&D en Ingeniería de Software y Sistemas de Información*
Departamento de Ciencias e Ingeniería de la Computación

UNIVERSIDAD NACIONAL DEL SUR

³*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)*

⁴*Comisión de Investigaciones Científicas de la provincia de Buenos Aires (CIC)*

Resumen

Esta línea de investigación se desarrolla en forma colaborativa entre docentes-investigadores de la Universidad Nacional del Comahue y de la Universidad Nacional del Sur, en el marco de proyectos de investigación financiados por las universidades antes mencionadas.

El objetivo general del trabajo de investigación es desarrollar una herramienta Web que permita la integración del soporte gráfico y el razonamiento automático en un ambiente de modelado conceptual. Se pretende trabajar en un arquitectura cliente-servidor y en la definición de un entorno gráfico con primitivas basadas en UML. De esta forma se podrán visualizar todas las deducciones relevantes modificando la apariencia del diagrama gráfico original y, dejando al usuario, la decisión de preservar o descartar dichos cambios.

Palabras Clave: Innovación en Sistemas de Software, Ingeniería de Software basada en Conocimiento, Lógicas Descriptivas, Ontologías.

Contexto

Este trabajo está parcialmente financiado por la Universidad Nacional del Comahue, en el marco del proyecto de investigación *Agentes Inteligentes en Ambientes Dinámicos (04/F006)*, por la Universidad Nacional del Sur a través del proyecto de investigación *Integración de Información y Servicios en la Web (24/N027)*, por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), en el contexto de una beca interna doctoral, y por el Consejo Interuniversitario Nacional (CIN), a través de una beca estímulo a las vocaciones científicas. Los proyectos de investigación tienen una duración de cuatro años y la beca doctoral una duración de 5 años, finalizando esta última en abril de 2019. La beca CIN tiene una duración de un año.

1. Introducción

Existe un incremento en la complejidad de los sistemas de información derivado de

nuevos conceptos, como por ejemplo, la Web Semántica [1], Big Data [2], e-government [3], etc. La calidad de estos sistemas está ampliamente determinada por el nivel conceptual, por lo tanto, el diseño de ontologías es clave para la posterior implementación y mantenimiento de dichos sistemas. En este contexto, los enfoques basados en lógicas para la representación de información y la adopción de técnicas de razonamiento automático deben asistir a los desarrolladores a obtener modelos conceptuales consistentes. Esta ingeniería ontológica necesita de metodologías y herramientas gráficas para la creación, edición y actualización de modelos que permitan establecer criterios de calidad claros y medibles.

Varias herramientas, entre las que podemos nombrar a NeOn toolkit [4], TopBrain Composer [5], Kaon [6] y Protégé [7], han sido desarrolladas para tratar con estos criterios. Sin embargo, aunque todas ellas permiten la edición y visualización de ontologías, la integración entre la parte visual y el razonamiento automático es muy débil. Por ejemplo, Protégé, ofrece un amplio conjunto de estructuras para modelado pero las inferencias obtenidas de razonadores OWL [8] externos están sólo limitadas a relaciones ISA en plug-ins gráficos.

En esta dirección, surge ICOM [9] como una nueva herramienta para diseñar y vincular múltiples modelos, facilitando su evolución en el tiempo. A diferencia de las herramientas previamente relevadas, la característica principal de ICOM es la combinación del diseño gráfico de ontologías con el soporte para razonamiento. ICOM fue desarrollada por la Free University of Bozen-Bolzano, y su objetivo principal fue demostrar la efectividad de los diagramas de clases para expresar ontologías y obtener nuevas conclusiones a partir de ellos. Si bien el lenguaje de modelado de ICOM permite expresar la mayoría de las características de EER [10] y UML [11], su comunicación con los razonadores es a través del protocolo DIG [12], que actualmente está siendo discontinuado por algunas herramientas. Por otra parte, el

soporte gráfico actual limita la visualización de características y restricciones al modelo.

En el ámbito de este trabajo, proponemos desarrollar una herramienta Web que permita la visualización de un modelo conceptual ampliando el soporte gráfico fuertemente integrado con el razonamiento automático y utilizando, como forma de comunicación, al protocolo OWLlink [13]. OWLlink es una evolución de DIG para el lenguaje OWL 2 [14]. La herramienta Web propuesta será utilizada, en el marco de nuestra investigación, como soporte para el desarrollo de nuevas y más expresivas metodologías asistidas por razonamiento, orientadas a cubrir algunas de las necesidades de la ingeniería ontológica identificadas anteriormente [15, 16].

La estructura del presente trabajo es la siguiente. En la sección 2 presentamos los objetivos de los proyectos de investigación en los que se enmarca este trabajo y describimos la línea de investigación actual. En la sección 3 indicamos algunos resultados obtenidos y trabajos futuros. Finalmente, comentamos aspectos referentes a la formación de recursos humanos en esta temática.

2. Línea de Investigación y Desarrollo

El proyecto de investigación *Agentes Inteligentes en Ambientes Dinámicos* tiene varios objetivos generales. Uno de ellos es el de *desarrollar conocimiento especializado en el área de Inteligencia Artificial*. En este sentido, se estudian técnicas de representación de conocimiento y razonamiento, métodos de planificación y tecnologías del lenguaje natural aplicadas al desarrollo de sistemas multiagentes.

Por otro lado, en el proyecto de investigación *Integración de Información y Servicios en la Web* se propone investigar y desarrollar metodologías y herramientas que favorezcan la interoperabilidad semántica de información y de servicios en la Web, fundamentados en los últimos avances en el área de

lenguajes de representación del conocimiento, ontologías y modelado conceptual.

Ambos proyectos confluyen en la línea de investigación de este trabajo, en la que se explora entre otros, sobre temas afines a la Representación del Conocimiento, las Lógicas Descriptivas [17], las Ontologías, la Ingeniería de Software basada en Conocimiento y la Ingeniería de Conocimiento. Particularmente, se ha escogido experimentar sobre metodologías que integren razonamiento con un *front-end* gráfico para dar soporte a la ingeniería de ontologías.

En esta línea de investigación se propone como principal objetivo desarrollar una herramienta Web cuya interfaz gráfica permita al usuario el diseño y la visualización de ontologías durante su evolución, asistido por técnicas de razonamiento automáticas. En este sentido, se ampliarán las características y restricciones que el usuario puede representar gráficamente, para lo que se analizarán técnicas de visualización y navegación [18], librerías gráficas y los lenguajes de modelado conceptual más utilizados en la actualidad, como EER, UML y ORM [19].

Para el diseño se propone una arquitectura cliente-servidor basada en tecnologías Web. Dicha arquitectura incluye una interfaz de usuario para la cual se definirán nuevas primitivas gráficas de modelado y se relevarán librerías disponibles para implementarlas. También, incluye un módulo de traducción a OWLlink cuyo objetivo es la representación de los modelos conceptuales en una lógica descriptiva, para su posterior procesamiento a cargo del razonador. Finalmente, se implementará un módulo generador de consultas para evaluar las propiedades de dichos modelos.

La arquitectura presentada anteriormente permite definir un proceso de verificación de los modelos conceptuales creados por el usuario. Cuando el razonamiento es invocado, la traducción OWLlink de los modelos es enviada al razonador junto con un conjunto de consultas generadas por el módulo pertinen-

te. A continuación, el sistema mostrará al usuario todas las deducciones relevantes modificando la apariencia del diagrama gráfico original y, dejando a éste, la decisión de preservar o descartar dichos cambios.

3. Resultados Obtenidos y Trabajo Futuro

Inicialmente, se diseñó una primer versión de la arquitectura cliente - servidor, que incluye entre otro los módulos de generación de consultas, librerías gráficas y un traductor para OWLlink. En este momento, la investigación se ha enfocado en el diseño del módulo gráfico.

Con el fin de que el desarrollo de la herramienta sea más ágil se ha relevado una serie de librerías gráficas, entre ellas Raphael¹, jsPlumb² y JointJS³. De este análisis, se determinó que la más adecuada era JointJS debido al nivel de complejidad en sus funcionalidades tanto para el desarrollador como para el usuario, su relación con Backbone⁴ y la posibilidad de crear y/o agregar plug-ins entre los cuales ya existen disponibles los que proveen las primitivas para EER y UML.

Asimismo, se ha creado un conjunto preliminar de primitivas gráficas basadas en los lenguajes de representación conceptual, como por ejemplo UML, a fin de hacer el uso de la herramienta más amigable.

Actualmente, se está trabajando en el módulo traductor para OWLlink y a futuro se trabajará sobre el generador de consultas a fin de concretar un primer prototipo de la herramienta.

Se espera poder validar el prototipo por medio de técnicas basadas en opinión, las que nos permitirá registrar las opiniones de los usuarios acerca de la navegabilidad de la herramienta y de la facilidad de aprendizaje y visualización de primitivas gráficas.

¹<http://raphaeljs.com/>

²<https://jsplumbtoolkit.com/>

³<http://www.jointjs.com/>

⁴<http://backbonejs.org>

4. Formación de Recursos Humanos

Uno de los autores de este trabajo está inscripto en el Doctorado en Ciencias de la Computación en la Universidad Nacional del Sur (beca interna doctoral CONICET).

En la Universidad Nacional del Comahue, Facultad de Informática, se otorgaron Becas CIN para estimular la vocación científica. Una de esas becas fue otorgada a uno de los autores de este trabajo, que está desarrollando su tesis de grado de la Licenciatura en Ciencias de la Computación en esta temática.

Referencias

- [1] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, May 2001.
- [2] Paul C. Zikopoulos, Chris Eaton, Drik deRoos, Thomas Deutsch, and George Lapis. *Understanding Big Data - Analytics for Enterprise Class Hadoop and Streaming Data*. 2012.
- [3] Chandra Misra. 1 defining e-government: A citizen-centric criteria-based approach, 2006.
- [4] P. Hasse, H. Lewen, R. Studer, and M. Erdmann. The NeOn Ontology Engineering Toolkit. 2008.
- [5] TopQuadrant. TopQuadrant — Products — TopBraid Composer, 2011.
- [6] B. Motik and R. Studer. KAON2—A Scalable Reasoning Tool for the Semantic Web. In *Proceedings of the 2nd ESWC'05*, 2005.
- [7] H. Knublauch, R. Fergerson, N. Noy, and M. Musen. The Protégé OWL plugin: An open development environment for semantic web applications. 2004.
- [8] W3C OWL Working Group. *OWL 2 Web Ontology Language: Document Overview*. W3C Recommendation, 27 October 2009. Available at <http://www.w3.org/TR/owl2-overview/>.
- [9] P. Fillottrani, E. Franconi, and S. Tessaris. The ICOM 3.0 intelligent conceptual modelling tool and methodology. *Semantic Web*, 2012.
- [10] Martin Gogolla. *Extended Entity-Relationship Model: Fundamentals and Pragmatics*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1994.
- [11] Grady Booch, James Rumbaugh, and Ivar Jacobson. *Unified Modeling Language User Guide*. Addison-Wesley Professional, 2005.
- [12] S. Bechhofer, R. Moller, and P. Crowther. The DIG Description Logic Interface. In *In Proc. of International Workshop on Description Logics (DL2003)*, 2003.
- [13] Thorsten Liebig, Marko Luther, Olaf Noppens, and Michael Wessel. OwlLink. *Semantic Web*, 2(1):23–32, 2011.
- [14] Markus Krötzsch. Owl 2 profiles: An introduction to lightweight ontology languages. In *Reasoning Web*, pages 112–183, 2012.
- [15] Germán Braun, Laura Cecchi, and Pablo Fillottrani. Integrating Graphical Support with Reasoning in a Methodology for Ontology Evolution. In *Proc. of the 9th Int. Workshop on Modular Ontologies WoMO 15 IJCAI 15*, CEUR Workshop Proceedings, 2015.
- [16] Germán Braun and Laura Cecchi. Extension Rules for Ontology Evolution within a Conceptual Modelling Tool. In *Proc. of the 1st Simposio Argentino de Ontologías y sus Aplicaciones SAOA 15 JAIIO 15*, CEUR Workshop Proceedings, 2015.

- [17] Diego Calvanese, Maurizio Lenzerini, and Daniele Nardi. Description logics for conceptual data modeling. In *Logics for Databases and Information Systems*, pages 229–263. Kluwer, 1998.
- [18] Colin Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.
- [19] Terry Halpin and Tony Morgan. *Information Modeling and Relational Databases*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2 edition, 2008.

Aplicación móvil con finalidad de ahorro sobre compras en supermercados

Ballardini Fabrizio, Cuevas Felipe, Fernández de Ullivarri Ramiro

Departamento Ingeniería Industrial, Departamento Ingeniería en Sist. de Información/ Facultad Regional
Córdoba/ Universidad Tecnológica Nacional

{[ballardinifabrizio](mailto:ballardinifabrizio@gmail.com), [felipecuevas17](mailto:felipecuevas17@gmail.com), [rfullivarri22](mailto:rfullivarri22@gmail.com)}@gmail.com

Resumen

En los últimos tiempos la continua inflación ha sido un factor dificultoso para la determinación de precios al momento de efectuar una compra en los supermercados. La incertidumbre de las familias con respecto a dicha temática llevó al desarrollo del siguiente proyecto: la sugerencia de compra más conveniente, comparando los montos de las mismas en distintos supermercados localizados dentro de un área preferencial, por medio de una aplicación móvil (app). Se intenta entonces, a través de bases de datos, juntar información útil y desarrollar algoritmos que ayuden a rastrear precios y actualizar los mismos. Además se desarrollaran modelos y gráficos estadísticos basados en el resultado de la información obtenida desde los usuarios, para poder hacer estudios de sondeo de mercado y brindarles dicho servicio de monitoreo a los supermercados (y posiblemente a las marcas).

Palabras clave: *Compra, Modelo Estadístico, Validación, Modelado Conceptual, Información actualizada, Sondeo.*

Contexto

El proyecto se lleva a cabo en el ámbito universitario, en colaboración con el Departamento de Ingeniería Industrial y el Departamento de Ingeniería en Sistemas de Información pertenecientes a la Universidad Tecnológica Nacional (UTN), Facultad Regional de Córdoba (FRC), en el marco del grupo de formación de emprendedores UTN Emprende y el programa de Mentoreo de Emprendedores ISI (Ingeniería en Sist. de Información) edición 2016, organizado por dicha institución.

Investigación, Desarrollo e Innovación

La app está dirigida a personas que por problemas de tiempo, dinero o localización tiene dificultad en la compra y en la decisión de compra en los diferentes supermercados.

La app puede dividirse en tres etapas que constituirían el ciclo de vida de la misma, para poder explicarla vamos a hacer esa división.

1°ETAPA:

-Objetivo: brindar a los usuarios un servicio virtual para el ahorro sobre la compra doméstica en supermercados.

Cliente directo: usuario

-Desarrollo: la app se va a encontrar en las diferentes plataformas de descarga. La descarga es gratuita y el usuario se registra por medio de Facebook o Google (esto facilita la experiencia del usuario que no debe estar registrándose y poniendo contraseñas con la posibilidad de olvidarla).

El usuario que accede, tiene una clasificación según rubros de los productos, y seleccionando el rubro entra a los distintos ítems del rubro (ej. Lácteos ----> yogur, leche...) y de ahí selecciona la marca de preferencia o alguno de los ítems. Y así hasta completar la compra que tiene pensado hacer.

El paso siguiente es la visualización de los precios de la compra total, en diferentes supermercados (a esto se puede agregarle un sistema de mapa de Google en el cual se sepa dónde queda el supermercado).

La app evidencia el más conveniente, y el usuario puede decidir acercarse o no a tal supermercado (posibilidad que el más barato quede muy lejos, entonces elige otra alternativa que le quede de paso).

La primera instancia de la app no tiene fin de generar ganancia, sino de ganar usuarios y “feedback” para poder usar estos datos en las siguientes etapas.

2°ETAPA:

Objetivo: facilitar y agilizar la experiencia de compra al usuario, y brindar a los supermercados estadísticas y tendencias de compra, generando (de común acuerdo con los negocios) ganancia adicional para el supermercado y un porcentaje para el mantenimiento de la aplicación.

-Cliente directo: Supermercado

-Cliente indirecto: Usuario

-Cliente posible: Marcas

-Desarrollo: en esta etapa se entra en contacto con los diferentes supermercados, presentando datos para poder acceder a un acuerdo con los mismos. A la app se agrega la posibilidad de armar la compra y retirarla en atención al cliente con una recarga en porcentaje sobre el monto total de la compra. El precio final que brinda la app al usuario aumenta su importancia ya que si el usuario estuviera interesado en pagar en efectivo sabría de antemano el monto a llevar al supermercado, en específico, a atención al cliente. En esta etapa también se puede entrar en contacto con las marcas para ofrecerle una "góndola virtual" dentro de la app. Se ofrece un espacio publicitario o estar entre los productos recomendados, por medio del pago de una cuota mensual por dicho servicio. De esta manera no se obstaculiza la inversión hecha por las marcas para posicionar sus productos en lugares estratégicos.

3°ETAPA:

Objetivo: Facilitar ulteriormente la experiencia de compra en supermercados a usuarios que por diferentes razones no pueden acceder a la compra directa en el dicho supermercado.

- Cliente directo: Supermercado
- Cliente indirecto: Usuario
- Cliente posible: Marcas

-Desarrollo: la segunda etapa se complementa con la tercera. En la tercera y última etapa el supermercado por medio de la app se acerca aún más al comprador por medio de un servicio de “delivery” (ya existente pero poco o casi nada difundido) en el cual se le lleva directamente a la casa (a horarios establecidos o no) la compra al usuario. El pago se efectúa en efectivo o tarjeta de débito o crédito directamente al encargado del “delivery” que posee “posnet” inalámbrico. La app sobre cada compra cobra una comisión al supermercado por el servicio (contacto usuario-supermercado).

Características y funcionalidades de la app

-Gráficos y estadísticas actualizadas automáticamente sobre tendencias de compra (producto y supermercado);

-Encuesta informal y divertida semanal, que familiarice al usuario con la aplicación, además de ser largamente útil para definir gustos y tendencias o sugerencias (a desarrolladores de la app, a supermercados y a marcas, todas reunidas en la misma base de datos);

-Interfaz amigable e intuitiva de fácil comprensión para personas no familiarizadas con la tecnología;

-Posibilidad de crear preferidos entre supermercados y marcas, acercando promociones e información útil al usuario

Con el siguiente proyecto se investigan las tendencias de compra de producto y

precios a lo largo del tiempo, determinando comportamientos dentro del mercado de productos de consumo masivo. Además se investiga e informa sobre la posición de supermercados en base a la ubicación preferencial. La Universidad UTN FRC aporta las herramientas para el desarrollo digital del proyecto, por medio de programadores y el acompañamiento en el desarrollo de la actividad comercial por medio de mentores y docentes emprendedores.

Resultados y Objetivos

Los objetivos del proyecto son:

- a) Brindar una aplicación que facilite y familiarice la experiencia de compra virtual del usuario;
- b) Brindar una herramienta de estudio y sondeo de las tendencias de compra y precios en el mercado;
- c) Efectuar un estudio económico basado en el objetivo anteriormente mencionado;
- d) Informar la localización de supermercados en el territorio de interés;
- e) Estudiar el mercado en cuanto a clientes, proveedores y posible competencia.

Formación de Recursos Humanos

El equipo del proyecto está compuesto por un grupo de estudiantes de Ingeniería Industrial y de Sistemas de Información. Se formará a los integrantes en cuanto a estudio y análisis de datos. Además se participará en cursos referentes al área específica que ocupa cada integrante para afianzar el conocimiento y obtener las herramientas para el desarrollo de su tarea dentro del proyecto. Se cuenta con un

mentor que es consultor de empresas perteneciente al grupo UTN Emprende que responde al nombre de Pablo Iachetti, brindado por la facultad, que se encarga de guiar y acompañar al grupo para el correcto desarrollo de la actividad además de brindar los instrumentos adecuados para la actividad. Además se tiene el respaldo de docentes de las cátedras de Ing. Industrial e Ing. En Sistemas de Información tales como el Ing. Oscar Medina y el Ing. Roberto Muñoz, con los cuales se colabora en diferentes actividades relacionadas al emprendedorismo para desarrollar las cualidades necesarias para la correcta realización del proyecto.

Referencia

[Levine 2006] Estadística para administración. Levine, David; Berenson, Mark L. Pearson Educación, 2006.

[Kotler 2015] Fundamentos de Marketing. Kotler, Philip. Addison-Wesley, 2015.

[Garibaldi 2014] Aprender Estadística, Aplicar Excel, Tomar Decisiones. Garibaldi, Carlos V. Libryco, 2014.

[Canvas] Canvas Business Model Generation.

[Maslow 1970] Maslow, Abraham. Motivation and Personality. Nueva York: Harper & Row, 1970.

[Deming 1982] Deming, W. Edwards. Quality, Productivity, and Competitive Position Cambridge, MA: Massachusetts Institute of Technology, 1982.

[Stanton 2015] Fundamentos de Marketing, Stanton, Etzel, Walker, McGraw Hill 11 Ed 14°, 2015

[Deming 2014] Calidad, productividad y competitividad: La salida de la crisis. Deming, W. Edwards. Diaz de Santos, 2014.

Linked Open Data para la Integración de Información Científica

Gaston Michelan¹ Germán Braun^{1,2,3} Laura Cecchi¹
Pablo Fillottrani^{2,4}

email: {gaston.michelan,german.braun,lcecchi}@fi.uncoma.edu.ar, prf@cs.uns.edu.ar

¹*Grupo de Investigación en Lenguajes e Inteligencia Artificial*
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

²*Laboratorio de I&D en Ingeniería de Software y Sistemas de Información*
Departamento de Ciencias e Ingeniería de la Computación
UNIVERSIDAD NACIONAL DEL SUR

³*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)*

⁴*Comisión de Investigaciones Científicas de la provincia de Buenos Aires (CIC)*

Resumen

Esta línea de investigación se desarrolla en forma colaborativa entre docentes-investigadores de la Universidad Nacional del Comahue y de la Universidad Nacional del Sur, en el marco de proyectos de investigación financiados por las universidades antes mencionadas.

El objetivo general del trabajo de investigación es relevar y analizar los nuevos lenguajes y tecnologías existentes, para generar y publicar datos abiertos e integrarlos con otras fuentes de datos disponibles en la Web. Asimismo, se proyecta identificar falencias de esta nueva infraestructura, proponer mejoras y evaluar su impacto en usuarios por medio de aplicaciones Web Semánticas. Se prevé utilizar como caso de estudio al grupo y laboratorio de afiliación de los autores.

Palabras Clave: Linked Open Data, Web Semántica, Ingeniería de Software.

Contexto

Este trabajo está parcialmente financiado por la Universidad Nacional del Comahue,

en el marco del proyecto de investigación *Agentes Inteligentes en Ambientes Dinámicos (04/F006)* y a través de una beca de Iniciación a la Investigación para alumnos; por la Universidad Nacional del Sur a través del proyecto de investigación *Integración de Información y Servicios en la Web (24/N027)*, y por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), en el contexto de una beca interna doctoral. Los proyectos de investigación tienen una duración de cuatro años y la beca doctoral una duración de 5 años, finalizando esta última en abril de 2019. La beca de Iniciación a la Investigación tiene una duración de un año.

1. Introducción

Del mismo modo que existen diversas lógicas para representar conocimiento con diferentes niveles de expresividad, en un lenguaje de marcado, también se debe considerar su capacidad para describir datos. En este contexto, HTML [1], por ejemplo, no es lo suficientemente expresivo para permitir que

ciertas entidades sean descriptas en un documento en la Web y que además puedan ser relacionadas entre ellas. Por otra parte, extraer información de fuentes tan heterogéneas disponibles en la Web es un problema complejo debido a que la cantidad de datos es vasta y se incrementa y actualiza constantemente. Una posible solución es estructurar los datos en un formato que permita acceder fácilmente a ellos, sin considerar el estado de las fuentes al extraer información, ya que estaremos siempre consultando la última actualización de los datos.

El objetivo principal de Linked Data [2, 3, 4], propuesto por Tim-Berners Lee, es publicar y conectar datos estructurados en la Web a través de un conjunto de buenas prácticas para tal fin. Así, nuevos documentos serán “entendibles” por las máquinas, tendrán un significado definido explícitamente y serán enlazados con otros, transformando la Web en una colección de tripletas RDF [5] referenciadas por URIs en los diferentes espacios de nombres. Esta capacidad para publicar y conectar datos propuesto por Linked Data es fundamental para la implementación de la Web Semántica [6, 7].

La iniciativa más importante en este campo es el Open Data Movement¹ cuyo objetivo, a través del “Linking Open Data Project”, es impulsar la publicación de datos disponibles para todos. En la actualidad existen varios conjuntos de datos abiertos, entre los que podemos nombrar DBpedia², WordNet³ y DBLP⁴, con el que estaremos trabajando en esta propuesta. La meta principal del proyecto es extender la Web publicando datos como RDF y enlazando dichos datos con otras fuentes también abiertas. Actualmente, esta iniciativa incluye más de 600 conjuntos de datos publicados, como muestra la última actualización del “LOD cloud diagram” en [8].

¹<https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

²<http://wiki.dbpedia.org/>

³<https://wordnet.princeton.edu/>

⁴<http://dblp.uni-trier.de/db/>

La comunidad científica no queda fuera de esta nueva filosofía y, en esta dirección, han surgido enfoques basados en Linked Open Data aplicados a las actividades de investigación [9, 10].

En el ámbito de este trabajo proponemos analizar diferentes tecnologías y formalismos que se requieren para poder ofrecer una fuente de datos enlazada y abierta, con información científica de grupos de investigación, como así también enfoques para su integración y herramientas para consumir estos datos que se beneficien de su forma estandarizada de representación y de la posibilidad de enlazar nuevas fuentes de datos en tiempo de ejecución. En particular, desarrollaremos una aplicación Web específica para el dominio académico combinando datos desde diferentes fuentes como, por ejemplo, DBLP, entre otras. Dicha aplicación permitirá visualizar la información actualizada asociada a los datos disponibles y realizar consultas en el lenguaje SPARQL [11] para RDF.

Se prevé utilizar como casos de estudio, al grupo de investigación GILIA del Departamento de Teoría de la Computación, de la Facultad de Informática, de la Universidad Nacional del Comahue y al LISSI, del Departamento de Ciencias e Ingeniería de la Computación, de la Universidad Nacional del Sur.

El desarrollo de una herramienta de estas características tendrá un gran impacto sobre nuestros equipos de investigación, ya que incorporarlos a la Web de datos, aumentará la visibilidad, fomentando la colaboración científica entre grupos interdisciplinarios.

La estructura del presente trabajo es la siguiente. En la sección 2 presentamos los objetivos de los proyectos de investigación en los que se enmarca este trabajo y describimos la línea de investigación, el problema que se estudia y los objetivos. En la sección 3 indicamos algunos resultados obtenidos y trabajos futuros. Finalmente, comentamos aspectos referentes a la formación de recursos humanos en esta temática.

2. Línea de Investigación y Desarrollo

El proyecto de investigación *Agentes Inteligentes en Ambientes Dinámicos* tiene varios objetivos generales. Uno de ellos es el de *desarrollar conocimiento especializado en el área de Inteligencia Artificial*. En este sentido, se estudian, entre otras, técnicas de representación de conocimiento y razonamiento aplicadas al desarrollo de agentes y a asistir el modelado conceptual.

Por otro lado, en el proyecto de investigación *Integración de Información y Servicios en la Web* se propone investigar y desarrollar metodologías y herramientas que favorezcan la interoperabilidad semántica de información y de servicios en la Web.

Estas líneas de investigación confluyen en el estudio de formalismos y tecnologías para cubrir las necesidades emergentes de compartir, actualizar e integrar el conocimiento de sistemas computacionales pre-existentes, elementos que se consideran fundamentales dadas las necesidades de interoperabilidad de aplicaciones, tanto a nivel de procesos como de datos. Particularmente, hemos escogido experimentar sobre los lenguajes y tecnologías existentes para generar datos abiertos y enlazados con otras fuentes de datos disponibles en la Web. Asimismo, se proyecta identificar falencias de esta nueva infraestructura, proponer mejoras y evaluar su impacto en usuarios por medio de una aplicación Web Semántica.

Como primer paso se trabajará en la estructuración de los datos de los grupos de investigación siguiendo los principios de Linked Data. Para satisfacer estos principios, se hará un relevamiento de las tecnologías disponibles para generar tripletas RDF a partir de nuestras bases de datos relacionales. Estas herramientas nos debe permitir la generación de URIs para identificar entidades unívocamente, cumpliendo así con el primer principio de Linked Data. Luego, para hacer públicos estos datos, se usará un servidor local que nos permitirá referenciar las URIs usando el pro-

toloco HTTP haciendo accesibles los datos desde fuentes externas, además de proveer la información requerida en el mismo formato RDF. Finalmente, insertaremos enlaces a otras URIs de datos RDF como, por ejemplo, [FacetedDBLP⁵](#), con información de publicaciones científicas. Además, incluiremos un motor para consultas SPARQL.

Si bien los datos, al estar disponibles en formato RDF, podrán ser consultados en el formato original con motores ad-hoc, es nuestra intención desarrollar una aplicación Web Semántica que permita visualizarlos y realizar consultas de un modo más amigable. La aplicación propuesta consumirá estos datos y habilitará a los usuarios a consultar la información requerida y navegar la Web de Datos a partir de su consulta inicial, en tiempo de ejecución, accediendo a la versión más reciente de ellos y sin considerar cómo y dónde están físicamente almacenados. De esta manera y, como un efecto directo de este trabajo, lograremos la incorporación de nuestros grupos a la Web de Datos fortaleciendo la integración, a nivel semántico, con otras fuentes de datos de interés.

3. Resultados Obtenidos y Trabajo Futuro

Inicialmente, se realizó un relevamiento de las tecnologías disponibles para nuestra arquitectura. Se analizaron las siguientes plataformas D2R [12], Sesame [13], Talis⁶, 4store [14], Jena [15] y Virtuoso [16], entre otras. De este análisis, se determinó que la plataforma más conveniente para un primer prototipo era D2R debido al nivel de complejidad de sus funcionalidades y la posibilidad de publicar datos de bases relacionales como triplas RDF.

Asimismo, a fin de lograr la integración entre los datos, se identificaron un conjunto de ontologías existentes para etiquetar nuestros

⁵http://dblp.l3s.de/?q=&newQuery=yes&resTableName=query_result3kKaKj

⁶<https://talis.com/>

datos y desambiguarlos. Como primera aproximación usaremos FOAF [17], para etiquetar personas, sus actividades y sus relaciones, y DC (Dublin Core) [18], para etiquetar recursos Web como, por ejemplo, artículos científicos junto con su título, autor, fecha, tipo y formato, entre otros.

Actualmente, se está trabajando en el desarrollo de un prototipo Web para visualizar y consultar estos datos enlazados. Como trabajo futuro, proponemos agregar más URIs de usuarios a otras bases de datos como lo son Bibsonomy⁷ y Geonames⁸, entre otras. Esto posibilitará el desarrollo de otras aplicaciones Web interdisciplinarias utilizando estos datos abiertos.

Hasta aquí, RDF nos ofrece un modelo basado en grafos para registro de datos, sin embargo, no nos permite agregarles semántica. Por lo tanto, también se espera poder definir un nivel más de abstracción, creando un vocabulario propio para describir entidades y relaciones del dominio. Luego, a partir de la ontología resultante, validar la consistencia de los datos y su relaciones implícitas utilizando técnicas de razonamiento basadas en lógica.

4. Formación de Recursos Humanos

Uno de los autores de este trabajo está inscripto en el Doctorado en Ciencias de la Computación en la Universidad Nacional del Sur (beca interna doctoral CONICET).

Otro de los autores, alumno avanzado de la carrera Licenciatura en Ciencias de la Computación, ha obtenido una Beca de Iniciación en la Investigación para Alumnos Universitarios de la Universidad Nacional del Comahue. Dicho becario realizará su tesis de grado en la temática de la línea de investigación presentada en el marco del GILIA.

Finalmente, nuestro programa de formación de recursos humanos incluye el dicta-

do de los cursos de posgrado “Ontologías y Web Semántica: Interoperabilidad Semántica de la Información”, en el marco del Doctorado en Ciencias de la Computación, UNS y “Gestión de contenido y tecnología de las redes sociales”, en el marco de la maestría en Bibliotecología y Ciencia de la Información⁹, UBA. En ambos casos el disertante es el Prof. Pablo Fillottrani, uno de los autores de este trabajo.

Referencias

- [1] W3C HTML Working Group. HTML, The Web’s Core Language, 2008. <https://www.w3.org/html/>, accedida en Marzo de 2016.
- [2] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked Data - The Story So Far. *Int. J. Semantic Web Inf. Syst.*, 2009.
- [3] Tom Heath and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 1st edition, 2011.
- [4] Liyang Yu. *A Developer’s Guide to the Semantic Web, Second Edition*. Springer, 2014.
- [5] Marcelo Arenas, Claudio Gutierrez, and Jorge Pérez. Foundations of rdf databases. In *Reasoning Web*, pages 158–204, 2009.
- [6] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, May 2001.
- [7] Tim Berners-Lee. Linked Data, 2006. <https://www.w3.org/DesignIssues/LinkedData.html>, accedida en Marzo de 2016.
- [8] Anja Jentzsch Max Schmachtenberg, Christian Bizer and Richard Cyganiak. Linking Open Data cloud diagram,

⁷<http://www.bibsonomy.org/>

⁸<http://www.geonames.org/>

⁹<http://maestriabiblio.blogspot.com.ar/>

2014. <http://lod-cloud.net/>, accedida en Marzo de 2016.
- [9] Tomi Kauppinen and Giovana Mirade Espindola. Linked Open Science—Communicating, Sharing and Evaluating Data, Methods and Results for Executable Papers. *Procedia Computer Science*, 2011. Proceedings of the International Conference on Computational Science, ICCS 2011.
- [10] Carsten Keßler and Tomi Kauppinen. Linked Open Data University of Muenster—Infrastructure and Applications. In *Demos ESWC*, Heraklion, Crete, Greece, May 2012.
- [11] W3C OWL Working Group. *SPARQL 1.1 Query Language*. W3C Recommendation, 21 March 2013. Available at <http://www.w3.org/TR/sparql11-query/>, accedida en Marzo de 2016.
- [12] Christian Bizer and Richard Cyganiak. D2R Server – Publishing Relational Databases on the Semantic Web. Poster at the 5th ISWC, 2006.
- [13] Jeen Broekstra, Arjohn Kampman, and Frank van Harmelen. Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. In *Proceedings of the First ISWC*, 2002.
- [14] 4store.org. 4store - Scalable RDF storage, 2012. <http://4store.org/>, accedida en Marzo de 2016.
- [15] Andy Seaborne. Jena, a Semantic Web Framework, November 2010. <http://wiki.apache.org/incubator/JenaProposal>, accedida en Marzo de 2016.
- [16] OpenLink Software. Virtuoso Universal Server, 2010. <http://virtuoso.openlinksw.com/>, accedida en Marzo de 2016.
- [17] Dan Brickley and Libby Miller. The Friend Of A Friend (FOAF) vocabulary specification, November 2007. <http://xmlns.com/foaf/spec/>, accedida en Marzo de 2016.
- [18] A. Powell, M. Nilsson, A. Naeve, and P. Johnston. Dublin Core Metadata Initiative - Abstract Model, 2005. White Paper.

La gestión de la información en abierto, vehículo importante para maximizar la visibilidad web

Jose Texier¹², Fernando Emmanuel Frati¹, Fernanda Carmona¹, Alberto Riba¹, Matías Pérez¹, Jusmeidy Zambrano¹²

¹Departamento de Ciencias Básicas y Tecnológicas, Universidad Nacional de Chilecito
9 de Julio 22, Chilecito, La Rioja, Argentina

{jtexier, fefrati, fbcarmona}@undec.edu.ar, albertoriba@gmail.com, mataguper20@gmail.com,
jzamban@unet.edu,ve

²Universidad Nacional Experimental del Táchira, Venezuela

Resumen

A mediados del siglo XX comienza la llamada "Era de la Información", definida por la investigación y el desarrollo en áreas como la microelectrónica, las telecomunicaciones y la informática. Esta era se caracteriza por reconocer a la información como un recurso valioso en todos los ámbitos, lo cual significa que hallar o generar la información a tiempo proporciona una gran ventaja competitiva. En este sentido, la UNdeC busca fortalecer su producción científica a través de una estrategia de gestión de la información de forma eficaz y eficiente, maximizando su visibilidad y garantizando una construcción correcta de los objetos de aprendizaje deseados. Por ello, con esta línea de I/D/I se busca concienciar a la comunidad universitaria respecto de algunos términos de uso frecuente en las áreas de visibilidad y gestión del conocimiento como lo son: los Repositorios Institucionales y el movimiento del Acceso Abierto, de manera tal que la generación de recursos educativos estén acordes con los lineamientos establecidos por la institución. El acceso al material académico y científico por parte de los

estudiantes, profesores y comunidad en general, con una garantía de acceso 24/7 sin restricciones técnicas, legales y de acceso, es un servicio indispensable para la garantía de la línea propuesta.

Palabras clave: producción científica, repositorios institucionales, acceso abierto, objetos de aprendizajes, visibilidad web

Contexto

Esta línea de I/D/I corresponde al diseño, desarrollo e implementación de proyectos que fortalecerán la investigación y las diferentes labores relacionadas con la gestión del conocimiento en cuanto a la visibilidad web y la preservación de la mayoría de la producción académica y científica de la UNdeC.

Actualmente se encuentra activo un PDTS (Proyectos de Desarrollo Tecnológico y Social) aprobado el año pasado y titulado "Repositorios digitales con contenidos orientados a las necesidades de escuelas rurales primarias y secundarias". La ejecución comenzará a partir del mes de abril de este año. También se encuentra activo el proyecto

“Red para la creación y publicación de objetos virtuales de aprendizajes de calidad en Repositorios Institucionales” correspondiente a la convocatoria “Redes Internacionales 9”, promovida por la Secretaría de Políticas Universitarias de la Nación.

Adicionalmente, están en evaluación un proyecto que fue presentado en la convocatoria 2013-2014 del programa “Financiamiento para el Estímulo y Desarrollo de la Investigación Científica y Tecnológica” de la Secretaría de Ciencia y Tecnología en la UNdeC (FiCyT - UNdeC) titulado “Fortalecimiento y visibilidad web de la producción científica de la UNdeC”. Sin embargo, el equipo de trabajo ya comenzó la ejecución del mismo con el levantamiento de la producción científica de la UNdeC a diciembre del 2015 y realizando un curso a nivel de postgrado llamado “Gestión de la Información Académica y Científica”.

En cuanto a la participación del grupo de trabajo en consorcios internacionales, se están desarrollando tres (3) propuestas para vincular la UNdeC con tres organizaciones internacionales con presencia al área de estudio: International Federation of Library Associations and Institutions (IFLA), Latin American and Caribbean Consortium of Engineering Institution (LACCEI) e Ibero american Science and Technology Education Consortium (ISTEC-BIREIDIAL).

Por otra parte se destaca, que la UNdeC cuenta con la estructura tecnológica y de RRHH necesarios para ejecutar los proyectos que surjan de la línea de investigación.

Introducción

Gestión/organización del conocimiento es un concepto que ha ganado cierta difusión, Barité y Dahlberg [1], [2]

señalan que cuando se habla de *organización del conocimiento* “es importante tener presente que (...) se hace referencia (...) [al] conocimiento socializado, compartido o comunicado, que además ha sido registrado”.

Por tanto, la información es la base para acceder al conocimiento. Se entiende al conocimiento como “una entidad abstracta que existe solamente en la mente de un ser humano en tanto sujeto cognoscente (es decir, es lo que yo sé), mientras que la información es el conocimiento comunicado, compartido o socializado” [3]. Esta diferenciación entre ambos conceptos (conocimiento e información) nos permiten comprender por qué es necesario que la información esté organizada para que el acceso y la apropiación [14] por parte de los sujetos les haga partícipes de la sociedad y los empodere como ciudadanos informados. Lo que en palabras de Barité y Dahlberg apunta a la socialización del conocimiento.

Sin embargo, esta situación se hace posible si esa información puede (o debe) registrarse en un soporte físico para convertirse en un documento o recurso bibliográfico con genuina utilidad por quien lo desee, por lo que los Repositorios Institucionales bajo la filosofía del Acceso Abierto de Objetos de Aprendizaje pueden convertirse en una de las vías más idóneas para lograr la visibilidad web que requiere la Institución.

Se entiende que los **Repositorios Institucionales (RI)** están constituidos por un conjunto de archivos digitales en representación de productos científicos y académicos que pueden ser accedidos por los usuarios [4]. En otras palabras, los RI se entienden como estructuras web interoperables que alojan recursos científicos, académicos y administrativos,

tanto físicos como digitales, descritos por medio de un conjunto de datos específicos (metadatos) [5]–[7]. Tienen como propósito recopilar, catalogar, gestionar, acceder, difundir y preservar la información [4], [8]. Vale la pena destacar que los RI son vías de comunicación científica, pero no son canales de publicación. Eso quiere decir que se deben seguir los mismos mecanismos de validación científica existentes hasta ahora a través de las revisiones por pares, pero los autores deben hacer énfasis en mostrar sus publicaciones y datos primarios de sus investigaciones por medio de las diversas formas que hoy ofrece la Internet.

Los RI se configuran dentro de la filosofía del **Acceso Abierto** (en inglés **Open Access - OA**). Esta filosofía tiene como fin asegurar el acceso libre y abierto a la producción científica, es decir, garantizar el acceso a través de Internet sin que los derechos del copyright sean una barrera [9], [10]. Los RI materializan el objetivo del OA porque la información que se deposita es una producción que tienen como propósito ser accesible, sin restricciones y preservada digitalmente como un bien común para la sociedad de hoy y del futuro.

El movimiento de acceso abierto a la información se basa en dos estrategias fundamentales, una a través de las revistas de acceso abierto y la otra por medio de los repositorios institucionales. En 1966, se conoce el lanzamiento de Educational Resources Information Center (ERIC), biblioteca digital especializada en educación, y de Medline, una base de datos bibliográfica de biomedicina producida por la National Library Medicine (NLM) de los Estados Unidos. Una de las voces líderes es Peter Suber [11], quien indica que existe una gran división en las publicaciones

científicas, una referida a aquellas que están disponibles gratuitamente en la Internet y, otra, en las cuales los lectores deben pagar para tener acceso a ellas, además, Suber enumera una serie de beneficios entorno al OA, destacándose que los artículos en acceso abierto han sido citados 50-300% más que artículos que no están en OA en una misma revista. Existen muchas declaraciones a favor del OA, pero se destacan tres: Berlin (2002), Bethesda (2003) y Budapest (2002). Estas declaraciones han permitido el surgimiento de otros movimientos bajo los mismos principios, a saber: Open Data, Open Knowledge o Data Sharing. En resumen, se pueden enumerar los siguientes objetivos del acceso abierto en los RI: (a) facilitar el acceso al conocimiento generado en la institución; (b) aumentar la visibilidad y el impacto de la producción intelectual de una institución; (c) dar mayor reconocimiento al autor; (d) promover generación de conocimiento a partir de la difusión y la vinculación con profesores, estudiantes e investigadores; y, (e) devolver a la sociedad parte del esfuerzo invertido (económico, humano, tecnológico) en la Universidad pública.

Como se ha mencionado los RI deben disponer también de **Objetos de Aprendizaje**, que son entidades generalmente entregadas a través de Internet y diseñadas con el fin de que sean utilizadas y reutilizadas en múltiples contextos educativos [12]. Para facilitar la reutilización de tales objetos de aprendizaje deben ser unidades independientes con un objetivo de aprendizaje claro y un contenido autónomo que no dependa de otros objetos [13]. Su unión puede generar nuevos objetos para ser utilizados en diversos ambientes educativos, definiéndose a estos no como simples elementos que pueden ser ensamblados,

sino como aquellos que por su naturaleza, permiten modificaciones en su misma composición interna para facilitar un correcto funcionamiento desde el punto de vista del diseño instruccional.

Sobre la base de lo expuesto, esta línea de investigación desarrollará proyectos que fortalezcan la difusión y gestión de la comunicación científica y académica de los becarios, docentes e investigadores de cualquier disciplina de la UNdeC. Esto se logrará mediante la socialización y sistematización de un conjunto de conocimientos prácticos que permitan alcanzar una visibilidad web que ayude a mejorar el posicionamiento de la UNdeC en el país y en el mundo. Todo esto bajo en contexto de la filosofía del Acceso Abierto.

Líneas de Investigación, Desarrollo e Innovación

- Repositorios institucionales y bibliotecas digitales.
- Objetos de aprendizajes.
- Gestión de la información y el conocimiento.
- Sistemas de información web y bases de datos.
- Índices bibliométricos.
- Interoperabilidad.
- Preservación digital.
- Recuperación de la información.
- Análisis semántico de la información.

Resultados y Objetivos

El desarrollo de esta nueva línea en la UNdeC permitirá cumplir con los siguientes objetivos:

- Difundir el concepto de conocimiento y la distinción de dato e información.
- Comprender los problemas y las tendencias asociadas a la producción abierta del conocimiento, el uso y la difusión, la educación abierta, la propiedad intelectual y licencias, y las desigualdades globales de comunicación centrales y transculturales.
- Conocer la mayoría de la producción científica de la UNdeC.
- Analizar los recursos educativos existentes que cumplan con las necesidades educativas de la UNdeC.
- Analizar la visibilidad web de la UNdeC.
- Fomentar la formación de habilidades y conocimientos relacionados con los RI.
- Desarrollar un mapa conceptual de la producción científica de la UNdeC.
- Entender y valorar el movimiento mundial de Acceso Abierto y señalar las vías de su materialización.
- Desarrollar las estrategias necesarias para optimizar la visibilidad científica de la UNdeC.

De igual forma, se espera conseguir los siguientes resultados:

- Conocimiento de la producción científica, lugar(es) donde está alojada y qué personal la realizó.
- Relevamiento de los diferentes recursos educativos producidos en la UNdeC
- Definición del posicionamiento web de la UNdeC.

- Elaboración de un mapa conceptual de la producción científica.

Diseño de un conjunto de estrategias para optimizar la producción científica de la UNdeC.

Formación de Recursos Humanos

El equipo de trabajo está formado por cuatro docentes de las carreras Ingeniería en Sistemas y Licenciatura en Sistemas de la UNdeC (acreditadas por CONEAU), dos doctores especializados en repositorios institucionales, bibliotecas digitales, desarrollo de software, cómputo paralelo y tecnología grid. Otra docente que está finalizando su doctorado y dos definiendo su tesis de Maestría en Informática. También participa un alumno avanzado de grado. En otras palabras, se cuenta con un recurso humano con habilidades y formación académica en las diversas áreas de la propuesta, asegurando la concreción de la línea, por ejemplo, uno es especialista en RI y OA, con investigación en esa área y otra está en formación (maestría) sobre el área de los Objetos de Aprendizaje. Adicionalmente, se destaca que dos están categorizados en el programa de incentivos.

Los integrantes son docentes de las asignaturas Programación I, Sistemas I, Arquitecturas Paralelas, Teoría de la Computación y Herramientas de Ingeniería de Software. Estas asignaturas contemplan la aprobación mediante la participación en proyectos de investigación, por lo que pueden surgir nuevos trabajos en esta línea.

Referencias

[1] M. Barité, *Diccionario de organización y representación del conocimiento*. Montevideo, 2000.

- [2] I. Dahlberg, “Knowledge organization: its scope and possibilities,” *Knowledge organization*, vol. 20, no. 4.
- [3] A. M. Martínez-Tamayo and J. C. Valdez, *Indicación y Clasificación en Bibliotecas*, Primera edición. Argentina: Alfagrama, 2009.
- [4] J. Texier, “Los repositorios institucionales y las bibliotecas digitales: una somera revisión bibliográfica y su relación en la educación superior,” presented at the 11th Latin American and Caribbean Conference for Engineering and Technology - 2013, Cancun, Mexico, 2013, p. 9.
- [5] C. A. Lynch, “Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age,” *ARL*., 2003. [Online]. Available: <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>. [Accessed: 28-Oct-2013].
- [6] J. Tramullas and P. Garrido, “Software libre para repositorios institucionales: propuestas para un modelo de evaluación de prestaciones,” *El Profesional de la Información*, vol. 15, no. 3, pp. 171–181, 2006.
- [7] H. Van de Sompel, S. Payette, J. Erickson, C. Lagoze, and S. Warner, “Rethinking Scholarly Communication,” *D-Lib Magazine*, vol. 10, no. 9, Sep. 2004.
- [8] M. De Giusti, N. Oviedo, A. Lira, A. Sobrado, J. Martinez, and A. Pinto, “SEDICI – Desafíos y experiencias en la vida de un repositorio digital,” *RENATA*, vol. 1, no. 2, pp. 16–33, Aug. 2011.
- [9] P. Suber, “Ensuring open access for publicly funded research,” *BMJ*, vol. 345, 2012.
- [10] D. Torres-Salinas, N. Robinson-García, and A. Cabezas-Clavijo,

- “Compartir los datos de investigación en ciencia: introducción al data sharing,” *Profesional de la Información*, vol. 21, no. 2, pp. 173–184, 2012.
- [11] P. Suber, “Open access, impact, and demand,” *BMJ*, vol. 330, no. 7500, pp. 1097–1098, May 2005.
- [12] D. Wiley, “Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy,” presented at the Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos, SPEDECE, 2001.
- [13] E. Morales, F. Garcia, A. Barron, A. Berlanga, and C. Lopez, “Propuesta de Evaluación de Objetos de Aprendizaje,” presented at the Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables (SPDECE), Barcelona, 2005.
- [14] Kalman J. (2001). *Saber lo que es la Letra*. Informe. México: DIE-CINVESTAV-IPN.

Diseño de sistema IoT de monitoreo y alarma para personas mayores

Sebastián Barillaro, Graciela De Luca, Waldo Valiente,
Esteban Carnuccio, Gerardo García, Mariano Volker,
Daniel Giulianelli, Nicanor Casas, Maximiliano Pérez

Departamento de Ingeniería e Investigaciones Tecnológicas /
Universidad Nacional de la Matanza/

Dirección: Florencio Varela 1703 - Código Postal: 1754

{sbarillaro, gdeluca, wvaliente, ecarnuccio, ggarcia, mvolker, dgiulian}@unlam.edu.ar
casas.nicanorb@gmail.com, maxiperezunlam@gmail.com

Resumen

En la actualidad los avances tecnológicos han permitido el desarrollo de productos y servicios que interactúan con personas y los objetos que los rodean. Estas tecnologías se orientan a simplificar y facilitar cuestiones de la vida diaria, sin embargo, en algunos casos, presentan una barrera. Especialmente para las personas de avanzada edad. Esta línea de investigación intenta encontrar una forma de derribar esas barreras tomando ventaja de las tecnologías actuales como sistemas embebidos con sensores de movimiento, sensores biométricos, conexión inalámbrica, geoposicionamiento, etc. Haciendo uso de estos dispositivos y de la computación en la nube, se busca brindar a las personas adultas mayores y de su entorno mayor comodidad, autonomía, independencia, reducción de costos en los cuidados, mejores controles de la salud y mayor celeridad en la respuesta ante emergencias.

Se busca crear una solución que monitoree la salud del usuario y la reporte a los cuidadores, familiares y médicos a través de internet durante las 24hs los 7 días de la semana. Además, emita alertas en los casos que el usuario requiera atención inmediata.

Palabras clave: computación en la nube, adultos mayores, geolocalización, monitoreo, IoT

Contexto

Nuestra Línea de Investigación es parte del proyecto “Sistema de monitoreo y alarma para personas adultos mayores ambulantes”, dependiente de la Unidad Académica del Departamento de Ingeniería e Investigaciones Tecnológicas perteneciente al programa de Investigaciones PROINCE de la Universidad Nacional de La Matanza, el cual es formado por docentes e investigadores de las carreras de ingeniería informática e ingeniería en electrónica. Este proyecto es continuación de los trabajos que viene realizando el grupo de investigación, en sistemas operativos, computación de alto rendimiento, entre otros, en el área de internet de las cosas (IoT).

Introducción

Los investigadores se han esforzado para crear, a través de la tecnología, una interfaz de humano a humano desde finales de la década de 1980. Lo que dio lugar a la creación de la disciplina computación ubicua, cuyo objetivo es

integrar la tecnología en el contexto de la vida cotidiana. Actualmente, tanto los teléfonos inteligentes como así también otros dispositivos portátiles están cambiando nuestro entorno, desplazando incluso a la computadora de escritorio en muchas prestaciones.

Mark Weiser, el creador de los primeros dispositivos utilizados en la computación ubicua (Ubicomp), define un entorno inteligente [1] como "El mundo físico que está ricamente y de forma invisible entretejido con sensores, actuadores, pantallas y elementos computacionales, integrado a la perfección en los objetos cotidianos de nuestras vidas, y conectado a través de una red continua".

La creación de Internet ha marcado un hito más importante en la consecución de la visión Ubicomp que permite a los dispositivos individuales comunicarse con cualquier otro dispositivo en el mundo. La interconexión de redes revela el potencial de una cantidad aparentemente interminable de recursos de computación distribuida y almacenamiento de datos de varios propietarios. Esta interconexión genera dos tecnologías críticas para el crecimiento de la infraestructura computacional en la nube o *Cloud-Computing* (CC) e Internet de las cosas (*IoT*, por sus siglas en Inglés).

Sumado a esto último, los avances y la convergencia de la tecnología de sistemas micro-electro-mecánicos (MEMS), las comunicaciones inalámbricas y la electrónica digital, produjeron el desarrollo de dispositivos en miniatura que tienen la capacidad de detectar, calcular y comunicar de forma inalámbrica de corto alcance. Estos dispositivos en miniatura llamados nodos de interconexión forman una red de sensores inalámbricos (WSN) y encuentran aplicaciones en vigilancia del medio ambiente, infraestructura de

vigilancia urbana, monitoreo de tráfico, comercio minorista, etc. El paradigma CC promete proporcionar servicios fiables entregados a través de los centros de datos de próxima generación, que se basan en tecnologías de almacenamiento virtualizados. Esta plataforma actúa como un receptor de datos de los sensores ubicuos; como una computadora para analizar e interpretar los datos; así como proporcionar al usuario una web fácil de entender, basada en la visualización. Los trabajos de detección y procesamiento ubicuos en el fondo, están ocultos para el usuario.

Para sacar el máximo provecho de la tecnología que Internet brinda, existe la necesidad de implementar a gran escala, infraestructura independiente de la plataforma, sensores inalámbricos en red que incluye la gestión y tratamiento de datos, actualización y análisis. También CC promete una alta fiabilidad, escalabilidad y autonomía para proporcionar acceso ubicuo, el descubrimiento de recursos dinámicos requeridos para la próxima generación de aplicaciones de IoT.

Líneas de Investigación, Desarrollo e Innovación

La hipótesis de desarrollo se basa en diseñar un prototipo de dispositivo portable equipado con sensores que permitirán conocer la ubicación de una persona en un ambiente interior y/o exterior, detectar cambios bruscos de movimientos, como caídas y cambios significativos en uno o más de sus signos vitales. Cualquiera de todas estas situaciones servirán de disparador automático de alarmas que serán recibidas por personas a cargo, generalmente un familiar del usuario. También se permitirá

el accionamiento manual de pedido de ayuda en caso de ser necesario [2].

Los datos recolectados por el dispositivo podrían ser transmitidos a un teléfono móvil o a una aplicación en la nube para un posterior procesamiento. Esta aplicación en la nube se encargará de llevar un monitoreo constante de los datos relevados al usuario, y alerta a los cuidadores inmediatamente al detectar anomalías que ameriten disparar una alarma. Todos los datos recolectados podrían ser útiles en futuras investigaciones o en control sanitario del paciente [3].

Se espera utilizar la combinación de sistemas embebidos conjuntamente con la transmisión segura de datos por internet, realizando su procesamiento en la nube. Lo que permitirá redundar en mayor ayuda y seguridad para las persona adultas mayores que lo requieran. Estas personas contarán con un cuidador o familiar a cargo que utilizará el sistema de monitoreo remoto provisto a un costo menor al de un cuidador presente [4].

En el mercado existen infinidad de sensores útiles para aplicaciones médicas, con costos accesibles. Encontramos tensiómetros, pulsómetros, glucómetros, acelerómetros, entre otros.

Además, la masificación de Internet y de los dispositivos móviles en la mayor parte de la población permite el diseño de un sistema de ayuda para personas de edad avanzada mediante el aprovechamiento del acceso a estas tecnologías.

Se han encontrado múltiples aplicaciones de salud para dispositivos móviles inteligentes. Existen empresas que instalan sensores de movimiento en todo el hogar conjuntamente con pastilleros inteligentes, tensiómetros conectados a la red, que entre otros servicios les permiten a los familiares del usuario saber la localización exacta de la persona que necesita ser monitoreada. También es

posible encontrar camas hospitalarias con sensores complejos para pacientes con internación domiciliaria, que realiza controles y en caso de ser necesario activa alarmas para permitir una atención médica urgente. Por ejemplo, BIODATA DEVICES [5].

También se reconoció la existencia de la necesidad de mejorar la calidad de vida de las personas mayores y de sus familiares mediante una tecnología que vaya en auxilio de sus necesidades en forma eficiente, amigable, segura y a un bajo costo, para poder ser adquirida por una gran parte de la población [6].

Resultados y Objetivos

Como la investigación se encuentra en etapas iniciales. Los objetivos planteados se centran en:

- Diseñar un sistema de monitoreo de salud, alerta y ubicación de personas adultas mayores ambulantes.

- Priorizar costos considerando como usuario final una población de bajos recursos, empleando tecnología existente en el mercado y adaptada para su fácil utilización.

- Permitir la transferencia y recolección de datos provistos por los sensores integrados, para que una persona receptora pueda monitorear parámetros vitales, la geolocalización y las eventuales caídas del usuario.

- Establecer los mecanismos de seguridad y confidencialidad de los datos.

- Desarrollar una aplicación en sistema embebido para recolectar, realizar la comunicación y transferencia de los datos que puedan ejecutar y conectarse con otras aplicaciones, en un teléfono inteligente y/o en una computadora de escritorio.

-Desarrollar un prototipo del dispositivo que llevará consigo la persona a monitorear.

-Generar documentación necesaria para que dicho sistema pueda ser reproducido y actualizado.

Para lograr estos objetivos se plantea una metodología de trabajo dividida en etapas que abarcan distintas áreas de conocimiento.

-Requisitos funcionales:

Del dispositivo, entre otros la ubicación si va a ser una pulsera o prenda en la ropa, la forma de alimentación y duración de batería, los sensores que se utilizarán, su mantenimiento, cómo transferirá los datos recolectados, entre otras cosas. Para el sistema CC, la forma en que guardarán los datos, seguridad de los mismos, la interfaz que brindara a múltiples dispositivos, soporte a distintos perfiles, reportes, entre otras cosas.

-Requisitos Técnicos y Tecnológicos:

Será necesario realizar un relevamiento de la tecnología disponible en cuanto a confiabilidad, costos, tamaño, alcances y prestaciones de los componentes que formarán el sistema portátil, teniendo en cuenta las limitaciones de las transmisiones de datos.

También será de suma importancia para las comunicaciones entre la persona monitoreada y el familiar o cuidador, ofrecer alternativas con distintos tipos de redes disponibles (tales como Wifi, Bluetooth, 3G, 4G), en caso de lentitud de la velocidad de transmisión o caída total de alguna, siempre debe existir una que la reemplace o permita el envío de datos.

-Diseño de solución:

Aplicar al prototipo portátil, como al sistema CC, con interfaces y mantenimiento.

Conclusiones

Para los ancianos, la idea de estar solo y tener una emergencia médica es una carga pesada con la que conviven diariamente. El sistema de monitoreo, geolocalización y alerta ofrece tranquilidad a esas personas y a las personas que se encargan de su cuidado.

Se debe tener en cuenta que una gran parte de las personas mayores tienen dificultades con el uso de la tecnología, por lo que el sistema debe ser extremadamente sencillo de utilizar, para no provocar el rechazo y brindar confianza al usuario y a quien lo monitorea.

Para la realización de este sistema se requiere observación, mejora de los prototipos, mediante pruebas en grupos de usuarios reales y realizar un análisis exhaustivo de las características de los componentes utilizados.

Formación de Recursos Humanos

La presente línea de investigación dentro del departamento de Ingeniería e Investigaciones Tecnológicas forma parte del trabajo que el Ingeniero Esteban Carnuccio se encuentra realizando para su maestría.

Dentro del grupo de investigación hay además dos investigadores en formación y se incorporó a un alumno que se encuentra finalizando su carrera, para realizar la iniciación en investigación.

Referencias

- [1] M. Weiser y R. Gold, «The origins of ubiquitous computing research at PARC in the late 1980s,» 1999.
- [2] J. Wherton, R. Procter, P. Sugarhood y S. Hinder, «Co-production in practice: how people with assisted living needs can help design and evolve technologies and services,» 2015.
- [3] A. Alonso, «Siembra de Datos,» *MIT Information Technology*, pp. 66-69, Diciembre 2015.
- [4] G. Atkinson y K. Karimi, «What the Internet of Things (IoT) Needs to Become Reality,» Texas;USA, 2014.
- [5] biodatadevices, «biodatadevices,» 2014. [En línea]. Available: <http://www.biodatadevices.com/index.php?lang=es>.
- [6] J. S. & S. K. A. Aquino, «La tecnología como apoyo para alertas y ubicación de grupos de interés prioritario.,» *Revista Científica y Tecnológica UPSE*, 2(2), pp. 1-6., 2015.

La Usabilidad a Través de Modelos Abstractos Empleando Desarrollo de Software Dirigido Por Modelos

Juan Carlos Moreno, Juan Pablo Fernandez Taurant, Marcelo Martín Marciszack

Departamento de Ingeniería en Sistemas / Universidad Tecnológica Nacional
(Facultad Regional Córdoba)

Maestro Lopez y Cruz Roja Argentina, Tel.: (0351)- 43320xx
{jmoreno33, jtaurant, marciszack} @gmail.com

Resumen

En el presente trabajo, se propone llevar a cabo un estudio para el planteo de una metodología, que permita capturar aspectos de usabilidad en forma temprana, partiendo de modelos conceptuales que se obtienen mediante la construcción de escenarios y el empleo de patrones. Este proceso formaría parte del ciclo de vida del software. La idea es incorporar aspectos de usabilidad en forma temprana, durante el proceso de construcción del software, partiendo de los modelos conceptuales y empleando patrones obtenidos de buenas prácticas del mercado. Para poder llevar a cabo esta propuesta se deberá realizar un análisis de las distintas metodologías actuales de diseño de sistemas y tomar en cuenta aquellas que permitan incorporar la usabilidad empleando escenarios y patrones en forma temprana. Se partirá desde el análisis del modelo de negocios hasta la obtención del modelo conceptual con la incorporación de patrones.

Palabras clave: Patrones de Usabilidad, Escenarios y Usabilidad, Modelado Conceptual con Criterios de Usabilidad

Contexto

La presente línea de Investigación se encuentra inserta dentro del proyecto de investigación “Validación de Requerimientos a través de Modelos Conceptuales” del Grupo de Investigación, Desarrollo y Transferencia

de Sistemas de Información, dependiente del Departamento Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, Facultad Regional Córdoba. Actualmente el grupo de investigación se encuentra trabajando en el “Proyecto de Implementación de Patrones en la validación de Modelos Conceptuales”.

Introducción

El desarrollo de sistemas de información web se ha transformado en un proceso que busca construir aplicaciones útiles y correctas para su uso. Esto implica tener mucho conocimiento técnico, metodológico, y de seguridad por parte del ingeniero de software. Uno de los objetivos de la Ingeniería de Software es construir aplicaciones de calidad, útiles a los usuarios finales, aplicando distintos métodos y principios [1]. La calidad de las aplicaciones web se mide muchas veces basándose en el sentido común y lógica de las experiencias de los desarrolladores [2].

Por lo general, en el proceso de construcción del software se hace énfasis en los aspectos de la arquitectura, la funcionalidad y la persistencia de cada proceso, no tratándose de forma adecuada la interacción y facilidad de uso. Por este motivo, el estudio de la interacción y de

la usabilidad del software web ha tomado relevancia.

El concepto de usabilidad ha sido definido por varias normas de Organizaciones Internacionales de Estándares de Calidad (ISO, IEEE) . En cada norma se presenta a la usabilidad como un atributo del software y está relacionado a la calidad del mismo.

La norma ISO/IEC 9126-1 [3], se considera a la usabilidad como un parámetro de calidad del software y es una de las características de calidad relevante del mismo. En esta norma se reconoce a la usabilidad como “la capacidad en que un producto de software puede ser entendido, aprendido y usado por determinados usuarios bajo ciertas condiciones en un contexto de uso específico”. Esta norma contempla la calidad interna, externa y en uso de un producto de software [4]. A su vez la usabilidad es descompuesta en subatributos como la facilidad de aprendizaje, la comprensión, operatividad y cumplimiento de la usabilidad [5].

La norma ISO 25000 (Square) [6] contempla a la usabilidad como un aspecto de calidad bajo dos puntos de vista distintos: uno que contempla a la usabilidad desde el punto de vista del software, como producto en sí mismo; y el otro punto de vista desde la usabilidad de uso, desde la perspectiva del usuario situado en un determinado contexto.

A través de los distintos estándares se definen distintos atributos de la usabilidad, que sirven para formular métricas para la evaluación del software.

Por lo general la usabilidad es considerada en etapas finales de la construcción del software. En esta etapa cualquier modificación afecta la arquitectura del sistema, puesto que las interfaces ya se encuentran diseñadas y el costo de cualquier modificación es alto [7], [8]. Una de las soluciones posibles a

este problema, es incluir el análisis de la usabilidad en etapas tempranas, durante la fase de elicitación de los requisitos. Por este motivo, es que se estudia el entorno de desarrollo de software dirigido por modelos (DSDM) [9], también denominado MDD en el campo de la Ingeniería de Software, puesto que se busca saber si se considera la elicitación de requisitos de usabilidad en etapas de desarrollo tempranas de la construcción del software. Esto afecta al sistema desde el punto de vista de su arquitectura. En DSDM se busca la construcción de un software a través de una serie de modelos conceptuales que son la representación del sistema de información. Partiendo de estos modelos se busca generar el código final del programa, aplicando una serie de transformaciones (de modelo a modelo y luego de modelo a código).

La incorporación del Desarrollo de Software Dirigido por Modelos (MDD) al proceso de desarrollo de software, fue realizada por Object Management Group (OMG), quien propuso una serie de especificaciones como estándares para formalizar este paradigma, independientemente de su implementación y que se denomina Model-Driven Architecture (MDA) (Miller y Mukerji 2003) [10]. Para ello utilizó otros estándares, como el Lenguaje de Modelado Unificado (UML) para modelado, Facilidad de Meta Objetos (MOF para metamodelos) y el lenguaje de intercambio de datos XML (adaptado y denominado XMI para transformaciones).

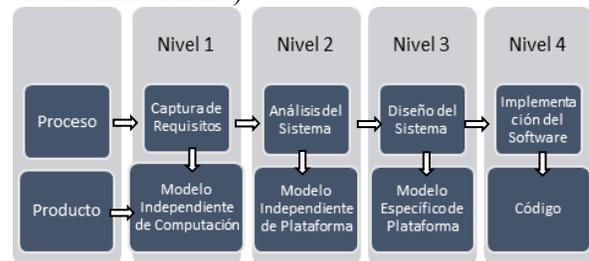


Figura 1. Model-Driven Architecture

Existen varios métodos de desarrollo de software de la Ingeniería web, que dan soporte al estándar MDD. Se pueden citar, como ejemplo, a los siguientes: OOHDM [11] [12], UWE [13] [14], OO-Method [15], OOH[16], OOWS[17], WebML[18].

Estos métodos, en su ciclo de vida, poseen un estilo de desarrollo del software en cascada. El desarrollo de los sistemas Web en estos métodos, se lleva a cabo mediante modelos que capturan distintas vistas del sistema: un modelo estructural (modela contenido y comportamiento), un modelo de navegación (modela acceso al contenido) y un modelo de presentación abstracto (modela cómo el contenido es mostrado). La evaluación de estos modelos puede proporcionar información, para evaluar la usabilidad en etapas tempranas del desarrollo antes de que cualquier línea de código sea generada. Se debe considerar el nivel de abstracción de los modelos a utilizar. Es relevante el nivel de expresividad de los modelos. Para poder evaluar las características de usabilidad.

Una vez obtenido un modelo conceptual (Casos de Usos, Diagramas de Interacción, etc), partiendo de un modelo de negocios sería interesante poder elaborar una serie de modelos abstractos, empleando teoría de grafos, donde se pueda aplicar la evaluación de Usabilidad y de accesibilidad simultáneamente. Esto es posible aplicando teoría de grafos a través de la evaluación de autómatas finitos, empleando de Redes de Petri Coloreadas y Autómatas Finitos Deterministas. Como consecuencia del Análisis Conceptual y a través de la determinación de los distintos objetos del sistema modelado, se debería analizar si es posible asociar los objetos obtenidos del modelo a patrones de interfaces y a instancias de tablas de bases de datos patrones, que posibilitasen incorporar los

campos estándares al patrón de interfaz seleccionado para trabajar y agregar los campos nuevos, en los casos necesarios.

Resultados y Objetivo

El objetivo general del presente trabajo será desarrollar una metodología, dentro del entorno Web, que permita incorporar aspectos de usabilidad partiendo de un modelo de negocios, y a través de transformaciones obtener modelos de escenarios con ciertos aspectos de usabilidad incorporados, los cuales mediante el empleo de patrones puedan instanciar interfaces finales, con un modelo conceptual integral y listo para su implementación. A través de dicho objetivo se busca analizar y evaluar el conocimiento actual de distintas metodologías, que dan soporte al desarrollo de software dirigido por modelos (DSDM) e implementen aspectos relacionados con la usabilidad en etapas tempranas del ciclo de vida de construcción del software de aplicaciones web.

En función de lo anteriormente expresado, se plantearon las siguientes metas:

- 1) Analizar distintas metodologías de diseño Web y seleccionar aquellas que consideran a la usabilidad en forma temprana, mediante el empleo de MDD.
- 2) Determinar de las metodologías de diseño Web seleccionadas, que herramientas emplean y entender la lógica que utilizan para incorporar aspectos de usabilidad en sus modelos conceptuales.
- 3) Explicar el procedimiento empleado para la construcción de las interfaces web, evaluando la factibilidad de incorporar atributos de usabilidad en forma temprana.

- 4) Hacer una reingeniería de procesos, que permita obtener una nueva metodología de usabilidad web temprano, mediante el empleo de transformaciones y patrones, que permita incorporar aspectos de usabilidad en forma eficiente y correcta.
- 5) Verificar la factibilidad de aplicar métricas, que permitan evaluar los resultados obtenidos.

Líneas de Investigación, Desarrollo e Innovación

Las Líneas de Investigación de Desarrollo involucradas son:

- a) Modelos de Negocios: se busca mediante el análisis del modelo de negocios la captura de los requerimientos de usabilidad en forma temprana.
- b) Aplicación del desarrollo de software dirigido por modelos, mediante el cual se llevan a cabo transformaciones para obtener un un modelo conceptual inicial.
- c) Obtención de los escenarios y diccionarios de datos, con criterios de usabilidad definidos.
- d) Obtención de modelos abstractos mediante transformaciones. Los modelos abstractos representan a las interfaces que son sometidas a evaluación utilizando grafos para determinar la presencia de usabilidad.
- e) Identificar los distintos tipos de patrones aplicables para la incorporación de usabilidad en forma temprana.

Objetivo y Resultados.

El objetivo general del proyecto de investigación es el siguiente:

- Establecer un marco teórico-metodológico para la incorporación de aspectos de usabilidad en forma temprana, a través del Desarrollo Dirigido por Modelos en una Metodología de Desarrollo Web, empleando modelos abstractos.

Asimismo, entre los resultados que se esperan a partir del presente trabajo, son:

- Establecer requisitos de usabilidad básicos, que debe poseer cualquier aplicación web.
- Describir métricas de usabilidad aplicables en metodologías de desarrollo web, que soportan el Desarrollo de Software Dirigido por Modelos.
- Verificar la existencia de herramientas o factibilidad de construcción de las mismas.

Formación de Recursos Humanos

Dentro del marco del proyecto de investigación actual, se doctoró el Mgter. Ing. Marcelo Marciszack, bajo la dirección del Dr. Manuel Pérez Cota, con su tesis de "Validación de especificaciones funcionales de esquemas conceptuales a través de abstracciones"; y están previstas la realización de las siguientes tesis y trabajos finales de carreras de Especialista en Ingeniería en Sistemas de Información U.T.N.- F.R.C. de la Ing. Claudia Castro titulado "Un modelo conceptual para la obtención del sistema de Información a partir del Sistema del Negocio"; trabajo de Tesis final dentro de la Maestría en Ingeniería en Sistemas de Información del tesista Esp. Ing. Juan Carlos Moreno cuyo tema es "Desarrollo de un marco teórico metodológico de técnicas de testing de usabilidad para verificar especificaciones no funcionales" cuyo director de Tesis es

el director del presente proyecto de investigación; y la postulación de Tesis de Doctorado en la U.T.N. F.R.C. de Oscar Carlos Medina bajo la dirección del Dr. Mario Groppo cuyo tema es "Metodología para implementación de patrones en buenas prácticas de e-Gobierno". Se incorporan tres becarios alumnos de investigación y un becario Graduado BINID, que colaboran en la investigación y construcción de una herramienta informática.

Referencias

- [1] Pressman R., "What a tangled Web we weave," *IEEE Software*, 2000.
- [2] Abrahao S., Condori-Fernandez N., Olsina L., and Pastor O., "Defining and validating metrics for navigational models," Australia, 2003.
- [3] Norma ISO/IEC ISO9126-1, "Software Engineering -Product Quality - Part 1," 2001.
- [4] Nigel Bevan, "Quality and usability: A new framework," *Achieving software product quality*, 1997.
- [5] Mario G. Piattini, Felix O. Garcia, and Ismael Caballero, "Calidad de Sistemas Informáticos," México, ISBN 978-970-15-1267-8, 2007.
- [6] ISO/IEC 25000, Software Engineering - Software Product Quality Requirements and Evaluation (SQuARE).
- [7] L Bass and B John, "Linking usability to software architecture patterns through general scenarios," *The journal of systems and software*, no. 66, pp. 187 - 197, 2003.
- [8] Eelke Folmer and Jan Bosh, "Architecting for usability: A survey. ," *Journal of Systems and Software*, pp. 61 - 78, 2004.
- [9] Stephen J Mellor, Kendall Scott, Axel Uhl, and Dirk Weise, *Model-Driven Architecture*. Berlin / Heidelberg: Springer, 2002.
- [10] (2003) MDA_Guide_Version1-0. [Online]. http://www.omg.org/mda/mda_files/MDA_Guide_Version1-0.pdf
- [11] G. Rossi and D. Schwabe, "Modeling and Implementing Web Applications using OOHDm," in *Web Engineering, Modeling and Implementing Web Applications.*: Springer, 2008, pp. 109-155.
- [12] S. Daniel, P. Rita de Almeida, and M. Isbela, "OOHDm-Web: an environment for implementation of hypermedia applications in the WWW," in *SIGWEB News 1.8, 2*, 1999, pp. 18-34.
- [13] Nora Koch and Martin Wirsing, *Software Engineering for Adaptive Hypermedia Applications*. München, Germany: Ludwig-Maximilians University of Munich, 2000.
- [14] N. Koch, A. Knapp, G. Zhang, and H. Baumeister, "UML-Based Web Engineering, An Approach Based On Standar.," in *Web Engineering, Modelling and Implementing Web Applications.*: Springer, 2008, pp. 157-191.
- [15] Oscar Pastor and Juan Carlos Molina, *Model-Driven Architecture in Practice: A Software Production Environment Based on Conceptual Modeling*, Inc. Secaucus, NJ, Springer-Verlag New York, Ed. Valencia, USA, 2007.
- [16] J. Gómez and C. Cachero, "OO-H Method: extending UML to model web interfaces.," in *In information Modeling For internet Applications.*, Hershey, PA.: Ed. IGI Publishing, , 2003, pp. 144-173.
- [17] P.V., Albert M., and Pastor O. Fons J.,.: LNCS. Springer, 2003, vol. 2813, pp. 232-245.
- [18] S. Ceri, P Fraternali, and A. Bongio, "Web Modeling Language (WebML): a modeling language for designing Web sites.," in *9th. World Wide Web Conference*, 2000, pp. 137-157.

Mejora de la precisión posicional utilizando receptores GPS de bajo costo

Alberto Eduardo Riba⁽¹⁾, Nelson Acosta^(2,3), Juan Manuel Toloza^(2,3), Fernando Emmanuel Frati⁽¹⁾, Jorge Tejada⁽¹⁾, Carlos Kornuta⁽²⁾

¹ Departamento de Básicas y Aplicadas, Universidad Nacional de Chilecito
9 de Julio 22, Chilecito, La Rioja, Argentina
{ariba, fefrati}@undec.edu.ar

² Universidad Nacional del Centro de la Provincia de Buenos Aires
General Pinto 399, Tandil, Buenos Aires, Argentina
{nacosta, jmtoloza}@exa.unicen.edu.ar, c_kornuta@hotmail.com

³ Universidad Nacional de Tres de Febrero
Mosconi 2736 - Sáenz Peña (B1674AHF), Buenos Aires, Argentina

Resumen

Los sistemas globales de navegación por satélite conocidos como GNSS son utilizados en innumerables ámbitos para georreferenciar objetos, el más antiguo y difundido es la tecnología NAVSTAR-GPS, pero existen otras implementadas como GLONASS y en vías de implementación como GALILEO y COMPASS que amplían esta oferta inicial.

En la actualidad una gran cantidad de dispositivos integran receptores GPS de bajo costo, que pueden ser utilizados en múltiples aplicaciones.

Para algunas de estas aplicaciones la precisión entregada por estos receptores no es suficiente.

Por este motivo la presente línea de I+D apunta al estudio, diseño y desarrollo de algoritmos, técnicas y métodos que permitan mejorar la precisión del posicionamiento entregada por este tipo de receptores GPS.

Los temas abordados son transversales a varias áreas, como algoritmos, lógica, programación, arquitectura de computadoras, sistemas de posicionamiento. Se espera que de esta

línea surjan futuras tesis de grado de las carreras relacionadas al área de la UNdeC, UNICEN y UNTREF.

Palabras clave:

Posicionamiento de precisión, GPS diferencial de bajo costo, Minimización de errores de sensores, Sistemas globales de posicionamiento, GNSS, GPS, DGPS

Contexto

Forma parte de esta línea de investigación el proyecto “Incremento de la precisión posicional relativa utilizando receptores GPS de bajo costo” presentado en la Secretaría de Ciencia y Tecnología convocatoria para estímulo y desarrollo de la investigación científica y tecnológica (FICyT - UNdeC - convocatoria 2013-2014), actualmente en desarrollo.

Cabe destacar que esta línea se inicia con el proyecto de tesis doctoral “Algoritmos y técnicas de tiempo real para el incremento de la precisión posicional relativa usando receptores GPS estándar”

del Dr. Juan Manuel Toloza becario CONICET.

Introducción

Durante años el posicionamiento de un objeto sobre la corteza terrestre ha sido objeto de innumerables investigaciones, desde los mercantes que usaban los cuerpos celestiales para ubicarse en su navegación por los océanos [1, 2], hasta los nuevos sistemas autónomos que, portando un micro-dispositivo embebido con distintos tipos de sensores, resuelven en cuestión de segundos la posición actual sobre la faz de la tierra para poder realizar su navegación [3].

Los GNSS (Global Navigation Satellite System) utilizan satélites para poder ubicar un receptor en tierra basado en técnicas de triangulación y medición de retardo de la señal [4]. Existen distintos sistemas GNSS implementados como NAVSTAR-GPS (Departamento de Defensa de EEUU) y GLONASS (Ministerio de Defensa de la Federación Rusa) y en vías de implementación como GALILEO (Unión Europea - GSA European GNSS Agency) y COMPASS (Gobierno Chino).

En nuestro país existe una gran cantidad de receptores de bajo costo que utilizan la tecnología NAVSTAR-GPS y existe una gran oferta de los mismos. Cuando se trabaja con este sistema utilizando los receptores mencionados, no se puede asegurar de obtener posiciones con una precisión mayor a los 15 metros el 95% de las veces [5].

Existe un gran número de aplicaciones que utilizan esta tecnología y pueden llevar a cabo su tareas sin inconvenientes en cuanto a la precisión obtenida, pero alguna áreas como la agricultura de precisión, la aeronavegación, la navegación marítima, los desarrollos

aeroespaciales, entre otras, necesitan una mayor precisión.

Varios desarrollos de empresas, gobiernos e instituciones afrontan a diario este desafío de encontrar nuevas técnicas para mejorar la precisión del posicionamiento, muchos de ellos con resultados exitosos y comprobables [6]. Pero estos desarrollos no alcanzan a la totalidad de los usuarios que los necesitan, en algunos casos las razones son económicas y en otros porque la región de residencia no se ve beneficiada por el servicio. Adquirir la infraestructura para montar un sistema de posicionamiento preciso, puede requerir de grandes inversiones. En otros casos puede implicar el abono de un servicio mensual. Aun así, y más allá de contar con la capacidad económica, en ciertas regiones no es posible acceder a las señales de corrección por características del terreno o por estar fuera del área de cobertura. Por este motivo es muy importante desarrollar soluciones tecnológicas que cubran estas necesidades.

Algunos requisitos operacionales para los que las constelaciones GPS y GLONASS no se elaboraron, se enuncian a continuación:

- **Exactitud:** diferencia entre la posición estimada y la real.
- **Integridad:** confianza sobre la información total proporcionada.
- **Continuidad:** funcionamiento sin interrupciones no programadas.
- **Disponibilidad:** es la parte del tiempo durante la cual el sistema presenta simultáneamente la exactitud, integridad y continuidad requeridas.

Para garantizar que los GNSS actuales cumplan con estos requisitos se requiere de diversos grados de aumentación. Se han diseñado y normalizado para superar

las limitaciones inherentes a los GPS tres sistemas de aumentación: el sistema basado en aeronave (Aircraft Based Augmentation System – ABAS), el basado en tierra (Ground Based Augmentation System - GBAS), y el basado en satélite (Satellite Based Augmentation System –SBAS). Para aplicaciones en tiempo real, las correcciones de los parámetros de cada satélite de las constelaciones GNSS existentes (GPS y GLONASS) deberán ser transmitidas a los usuarios a través de equipos de radio VHF (GBAS) o si se requiere una amplia cobertura a través de satélites geostacionarios que emitan pseudocódigos con información de corrección (SBAS) [7].

Esta línea de investigación pretende reducir la brecha existente entre estos desarrollos y los usuarios finales que necesitan realizar tareas con mayor precisión posicional que la entregada originalmente por un sistema global de navegación satelital (GNSS) como es el NAVSTAR-GPS, utilizando receptores GPS de bajo costo.

Líneas de Investigación, Desarrollo e Innovación

- Posicionamiento de precisión
- Sistemas de Tiempo Real
- Protocolo NMEA
- Geometría de los satélites
- Posicionamiento diferencial de bajo costo
- Minimización de errores de posicionamiento

Resultados y Objetivos

Resultados

Como resultados se puede mencionar el trabajo de tesis doctoral de uno de los integrantes, que desarrolló una herramienta donde se implementa un conjunto de técnicas y algoritmos para el tratamiento de información que proviene de los receptores para incrementar la exactitud del posicionamiento.

Esta herramienta es totalmente configurable y posee portabilidad de manera que funcione en cualquier región donde no se cuenta con servicios de aumentación para mejorar la precisión.

Durante 2015 se adquirió el equipamiento necesario para realizar nuevos experimentos que permitan analizar la geometría de los satélites. Actualmente el equipo está trabajando en el análisis y desarrollo de nuevas técnicas y algoritmos que anexadas a las ya propuesta permitan mejorar la precisión posicional en base a los datos obtenidos.

Además se logró fortalecer el grupo de trabajo, como así también incentivar la producción científica en la UNdeC.

Objetivo General

Desarrollar un conjunto de técnicas y algoritmos de mejora a la precisión del posicionamiento utilizando receptores GPS estándar de bajo costo en un prototipo de GPS diferencial.

Objetivos Específicos

Analizar técnicas y algoritmos de mejora de la precisión del posicionamiento para receptores GPS estándar. Analizar y diseñar técnicas para el análisis geométrico de la posición de varios satélites. Estudiar el impacto que tiene el factor de dispersión espacial de los satélites en el error de posicionamiento de un GPS de bajo costo. Analizar técnicas y algoritmos de la integridad de los datos recibidos por el

dispositivo. Analizar técnicas de
aumentación de la precisión.

Formación de Recursos Humanos

El equipo de investigación que lleva a
cabo esta línea de trabajo en
posicionamiento está compuesto por 5
docentes y 3 alumnos de grado.

De los docentes: 1 es posdoctorado en
Informática, 2 son doctores en
Informática; un maestrando que
presentará su tesis en la Universidad
Nacional de San Juan; un doctorando que
presentará su tesis en la Universidad
Nacional del Centro de la Provincia de
Buenos Aires. Cabe destacar que el
doctorando es becario de CONICET. Los
tres alumnos de grado se encuentran
desarrollando dos tesinas finales de grado
en el marco de este proyecto.

Referencias

[1] Rao (2010) Global Navigation Satellite
Systems. Tata McGraw-Hill Education,
478 pp.

[2] Misra P. & Enge P. (2010) Global
Positioning System: Signals,
Measurements, and Performance. New
York, Ganhga-Jamuna Press, 590 pp.

[3] Enrique David Martí, David Martín,
Jesús García, Arturo de la Escalera, José
Manuel Molina and José María Armingol.
“Context-Aided Sensor Fusion for
Enhanced Urban Navigation”. Open
Access Sensors, Article.

[4] Gleason S. & Gebre-Egziabher D.
(2009) Gns Applications and Methods.
Artech House, 508 pp.

[5] Zandbergen P. A. & Arnold L. L.
(2011) Positional accuracy of the Wide
Area Augmentation System in consumer-
grade GPS units. Computers &
Geosciences Volume 37 Issue 7, Elsevier,
pp. 883-892.

[6] Tolosa Juan Manuel. (2012)
“Algoritmos y técnicas de tiempo real
para el incremento de la precisión
posicional relativa usando receptores GPS
estándar”. SEDICI, Universidad Nacional
de La Plata.

[7] Alberto Riba, (2015) “Incremento de
la precisión posicional relativa utilizando
receptores GPS de bajo costo,” in IV
Jornadas Científicas de Estudiantes
Investigadores (IV-JCEI), Universidad
Nacional de Chilecito, Argentina.

Comunicación alternativa y aumentativa para potenciar la autonomía personal y la calidad de vida de las personas con discapacidades severas

Javier Díaz, Laura Fava, Ivana Harari, Fernando Martinez, Miguel Telechea
LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas.
Facultad de Informática. Universidad Nacional de La Plata
Calle 50 esq. 120, 2do Piso. Tel: +54 221 4223528
{jdiaz, lfava, iharari}@info.unlp.edu.ar, {tellecheamiguel,
fernandomartinez@gmail.com}

Resumen

El lenguaje oral es una de las formas básicas para comunicarnos con el entorno y muchas veces nos sirve para modificarlo en función de nuestras necesidades y deseos. Si una persona no puede comunicarse, no podrá tomar decisiones que afecten a su vida, lo que originará dependencia y aislamiento.

La comunicación aumentativa y alternativa es la combinación de métodos y estrategias de comunicación utilizados por personas con determinadas discapacidades para una comunicación hablada o escrita. Existen múltiples sistemas, dispositivos y recursos que favorecen la comunicación cuando existen dificultades para hacerlo, pero a veces no alcanza cuando la discapacidad es severa. En este proyecto se propone analizar sistemas de software de comunicación aumentativa y alternativa y evaluar su integración con dispositivos especiales, en particular aquellos relacionados con interfaces cerebro-computadoras, para potenciar la autonomía personal y la calidad de vida de personas con discapacidad agudas.

Palabras claves: comunicación aumentativa, comunicación alternativa, CAA, interfaz cerebro computadora, Brain Computer Interface (BCI), Emotiv EPOC, Emptiv Insight.

Contexto

La Facultad de Informática, a través de la Dirección de Accesibilidad y del Laboratorio de Investigación de Nuevas Tecnologías Informáticas LINTI, ha llevado a cabo en los últimos años, líneas de acción concretas que estrechan el vínculo Facultad-Sociedad, atendiendo las demandas de los sectores más vulnerables de la comunidad, como lo son las personas con discapacidad. A través de varios cursos de formación abierto a todo público como el de Accesibilidad Web, el de TICs para personas con discapacidad; de jornadas y conferencias como la de una Facultad Inclusiva, la de Experiencias y Casos de Aplicación de desarrollos accesibles, las hackatones de 24 hs de desarrollo por la discapacidad; como también, a través de la dirección de proyectos sobre inclusión social y accesibilidad como el Zapato Háptico y Juegos Serios para Autistas, entre otros; hace que la Facultad de Informática se

constituya como una entidad de referencia respecto a cuestiones de tecnología y discapacidad.

Durante estos años de comunicación y de interacción continua con la comunidad, se elaboraron proyectos que permiten intervenir ante las necesidades reales de las personas con discapacidad, como es el caso de discapacidades severas de motricidad y de habla.

El proyecto que se describe en este artículo realiza un abordaje de esta problemática mediante actividades de investigación, innovación y desarrollo y de búsqueda permanente de soluciones tecnológicas que mejoren la calidad de vida de las personas que atraviesan este grado de discapacidad.

El mismo se desarrolla en el Laboratorio de Investigación en Nuevas Tecnologías Informáticas, LINTI de la Facultad de Informática de la UNLP y está enmarcado en los proyectos 11-F014 *Innovación en TICs para el desarrollo de aplicaciones en educación, inclusión, gobierno y salud* finalizado en 2015 y en el nuevo proyecto *Internet del futuro: Ciudades digitales inclusivas, innovadoras y sustentables, IoT, ciberseguridad y espacios de aprendizaje del futuro*, ambos acreditados en el marco del Programa de Incentivos, bajo la dirección del Lic. Javier Díaz.

Introducción

Existen múltiples dispositivos, sistemas y recursos que favorecen la interacción comunicativa cuando existen dificultades para que pueda llevarse a cabo. Dentro de estos productos podemos encontrar sistemas de comunicación aumentativos y alternativos (CAA). Los sistemas aumentativos de comunicación, son aquellos que complementan el lenguaje oral cuando, por sí sólo, no es suficiente para entablar una comunicación

efectiva con el entorno [Ref. 1]. Los sistemas alternativos de comunicación, son aquellos que sustituyen al lenguaje oral cuando éste no es comprensible o está ausente [Ref. 2].

Si bien estas propuestas importan una solución o alternativa a para personas con determinadas discapacidades, las personas que padecen cuadriplejía, esto es, parálisis total o parcial de sus extremidades superiores e inferiores, se encuentran limitados para comunicarse con soluciones basadas solo en CAA, lo que provoca inevitablemente una fuerte dependencia y aislamiento.

Afortunadamente, en los últimos años han surgido alternativas como las interfaces cerebro computadora (BCI por sus siglas en inglés) que prometen una esperanza para mejorar la situación de estos pacientes. Las BCI son dispositivos que pueden leer las señales eléctricas del cerebro humano y traducirlas en señales de control para un dispositivo. La tecnología BCI provee un canal de comunicación alternativo que permite a los usuarios enviar información a un dispositivo a través de su pensamiento. Si bien fueron diseñadas principalmente para permitir la comunicación de pacientes con discapacidad neuromuscular severa, los trabajos multidisciplinarios, el progreso de la neurociencia cognitiva, el procesamiento de señales y reconocimiento de patrones, ha inspirado el uso de BCI como una nueva modalidad para interacciones cerebro-computadora [3, 4].

Este artículo presenta una línea de investigación cuyo objetivo central es analizar herramientas alternativas y aumentativas y evaluar su integración con dispositivos cerebro-computadora para proveer una solución a pacientes con discapacidades severas.

Líneas de Investigación,

Desarrollo e Innovación

Las líneas de investigación, desarrollo e innovación que se llevan a cabo en este proyecto se basan en dos ejes principales: uno vinculado a la investigación de softwares aumentativos y alternativos existentes y otra a la investigación de dispositivos que permitan aumentar la capacidad de los mismos.

En la actualidad existen muchos sistemas aumentativos y alternativos y nuevos dispositivos que facilitan el desarrollo de nuevo productos. A continuación sintetizamos algunos de los softwares evaluados y seleccionados para profundizar su estudio e integrarlos con dispositivos BCI.

- **ACAT** (Assistive Context-Aware Toolkit) es una plataforma de código abierto desarrollada en los laboratorios de Intel para permitir que las personas con enfermedades de las neuronales, motoras y otras discapacidades puedan tener acceso completo a las computadoras a través de interfaces muy limitados adecuados para su condición. Más específicamente, ACAT permite a los usuarios comunicarse fácilmente con otros a través de la simulación de teclado, la predicción de palabras y la síntesis de voz. Los usuarios pueden realizar una serie de tareas tales como la edición, gestión de documentos, navegar por la Web y acceder a mensajes de correo electrónico. ACAT fue originalmente desarrollado por investigadores de Intel para el profesor Stephen Hawking, a través de un proceso de diseño muy iterativo en el transcurso de tres años. ACAT ha sido recientemente liberado bajo la licencia de código abierto de Apache v2, para que cualquier persona lo

pueda descargar y usar sin costo [Ref.5].

La liberación de código de ACAT representa un gran avance en el desarrollo de herramientas de accesibilidad, donde millones de personas pueden ser beneficiadas, ya que cualquier desarrollador en el mundo podrá experimentar y crear nuevos sistemas con esta base.

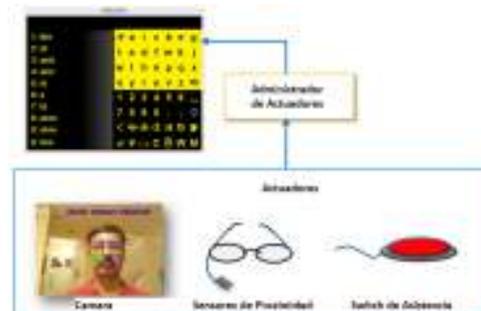


Figura 1: Componentes principales del ACAT

La Figura 1 muestra las componentes principales del software ACAT: una cámara web, sensores de proximidad y un dispositivo de asistencia.

- **Plaphoons** es un programa de comunicación para personas con discapacidad motriz, que en principio fue pensado como sistema de comunicación simbólica, tanto desde el punto de vista de ser utilizado de forma directa, como para la edición de plafones de comunicación en formato papel. Ha resultado ser una herramienta muy útil para el aprendizaje de la lectoescritura a dichas personas, para al final, servir de sistema de escritura en cualquier procesador de texto y también como sistema de control a diversos programas. La finalidad principal de este programa es dar más independencia a estas personas permitiendo que construyan sus mensajes, sugerencias o ideas de forma totalmente independiente. Este

software puede descargarse de manera gratuita, pero no es de código abierto [Ref. 6].



Figura 2: Interfaz de usuario de Plaphoons

La figura 2 muestra uno de los varios tableros de comunicación que provee del Plaphoons.

- **HADASoft** es un software integrado por distintos módulos que, combinados, hacen posible la escritura y la reproducción del habla a personas con discapacidad. Este software hace uso de un teclado virtual unido a una cámara web, para permitir que los usuarios con discapacidad física manejen el cursor con la mirada y un sintetizador que reproduce en audio las palabras seleccionadas. La cámara web está especialmente programada para hacer un seguimiento de los movimientos de la cabeza, en especial de la mirada, de ahí su nombre "head mouse". Este software puede descargarse de manera gratuita desde la web, pero no es de código abierto. [Ref. 7]. HeadMouse es un software desarrollado por el equipo del Neural Information Processing Group, de la Universidad Eötvös de Budapest, Hungría. El software es muy estable y la configuración sencilla. La Figura 3 muestra la ventana de configuración de este mouse adaptativo.

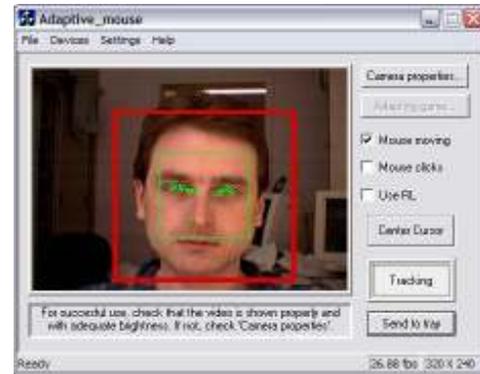


Figura 3: Pantalla de configuración del mouse

Además de profundizarse el estudio de estos softwares seleccionados, se evaluará su integración con interfaces cerebro-computadoras para permitir su uso a personas con discapacidades severas. Para ello se investigarán interfaces cerebro computadoras, en particular, Emotiv EPOC que se describe debajo.

- **Emotiv EOC** es una BCI inalámbrica, multicanal, de alta resolución. Fue desarrollada para permitir la interacción humano-computadora a través de la detección de pensamientos, sentimientos o expresiones en tiempo real. A diferencia de otros dispositivos BCI existentes como el Mindball que posee 3 electrodos, el MindSet que provee solo 1 electrodo o el Neural Impulse Actuator que tiene 3 electrodos, el Emotiv EPOC cuenta con 14 electrodos y un sensor giroscópico.



Figura 4: Emotiv EPOC/Emotiv INSIGHT

La Figura 4 muestra a la izquierda una imagen del Emotiv EPOC descrito y a la derecha una nueva propuesta de Emotiv, llamado Emotiv INSIGHT, con 5

sensores pero de mejor calidad y facilidad de uso.

La electroencefalografía (EEG) es el registro de la actividad eléctrica a lo largo del cuero cabelludo a través de la medición de las fluctuaciones de tensión que acompañan a la actividad de la neurotransmisión en el cerebro. Los electrodos se conectan en un dispositivo similar a una tapa, como se muestra en la Figura 4. Tiene algunas ventajas de usabilidad únicas sobre otros tipos de grabación de la señal cerebral ya que es fácil de usar, portátil y de bajo costo, sin embargo su relación señal a ruido representan una limitación en comparación con otros métodos [8]. Es oportuno observar que el EPOC para usar el casco se debe acondicionar al paciente (humedecer cada uno de los 16 conectores y colocarlos en el lugar correcto).

Además del dispositivo físico o casco, el EMOTIV provee un SDK para desarrollar aplicaciones sobre Windows, Linux, MAC, iOS y Android y algunas herramientas para probar el casco.

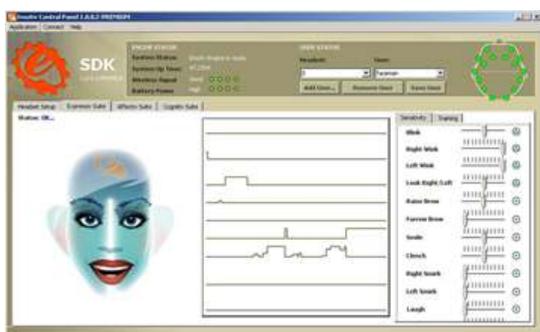


Figura 5: Panel de Control del Emotiv EPOC

La Figura 5 muestra el Panel de Control del Emotiv EPOC, versión desarrollador, donde puede observarse la visualización del reconocimiento de expresiones faciales adicionales como surco de la frente, sonrisa hacia derecha, sonrisa hacia izquierda y risa. También introduce la capacidad de entrenar el

sistema con su propia cara. Por último, en el sitio de EMOTIV, se encuentran algunas aplicaciones para descargar y probar, algunas pagas, otras gratuitas, pero ninguna de código abierto [9].

Resultados y Objetivos

Como ya ha mencionado, en el LINTI se han desarrollado muchas herramientas que permiten mejorar la calidad de personas con discapacidad como el zapato háptico, Juegos Serios para autistas, aplicaciones para disminuidos visuales, etc. Este proyecto continúa esa línea de trabajo con el análisis de diferentes aplicaciones aumentativas y alternativas y el estudio de interfaces cerebro computadoras con el objetivo de integrarlas con las aplicaciones CAA evaluadas.

El proyecto presentado en este artículo está en una etapa intermedia, razón por la cual los resultados representan las primeras pruebas de tecnología y análisis parcial de las aplicaciones.

Los resultados alcanzados en esta etapa son:

- Análisis de patologías severas donde los pacientes perdieron la movilidad: pérdida de los sentidos y de la conciencia.
- Investigación y selección de diferentes softwares para comunicación aumentativa y alternativa, teniendo en cuenta sus funcionalidades provistas y posibles customizaciones.
- Estudio, prueba de dispositivos especiales, en particular del Emotiv EPOC y su SDK. Las primeras pruebas de estos dispositivos nos han permitido verificar que las interfaces cerebro computadora han resultado ser mucho más lentas, más ruidosas y más propensas a errores en comparación con otros dispositivos

de entrada, además de las dificultades que involucra su uso.

- Análisis, diseño e implementación de un primer prototipo web simple para determinar la forma óptima de comunicación y la respuesta del paciente. Esta aplicación está comenzando a ser utilizada por un equipo de especialistas y con un paciente con discapacidad severa, donde se está trabajando su rehabilitación.

Los próximos objetivos están relacionados con:

- Estudiar la confiabilidad de las señales entregadas por la interfaz cerebro-computadora mediante un análisis del comportamiento de las mismas y así determinar el grado de controlabilidad que estas señales ofrecen para el manejo de la aplicación.
- Integrar softwares aumentativos y alternativos con interfaces cerebro-computadora.
- Analizar el impacto en el paciente, el nivel de independencia logrado y el grado de mejora de su calidad de vida.

Formación de Recursos Humanos

El equipo de trabajo de la línea de I+D+i presentada en este artículo se encuentra formado por docentes investigadores categorizados del LINTI y alumnos avanzados de las carreras de Lic. en Informática y Lic. en Sistemas de la Facultad de Informática.

Basada en las líneas de investigación de este proyecto se está desarrollando una tesina de grado.

Referencias

[1] Desney S. Tan y Anton Nijholt (2010). *Brain-Computer Interfaces, Applying our minds to Human-Computer Interaction*. New York: Springer. ISBN 978-1-84996-272-8

[2] Abadín, D, Delgado Santos C. y Cerrato A. (2010). *Comunicación Aumentativa y Alternativa, guía de referencia*. Madrid: CEAPAT

[3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control", *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767-791, Junio 2002.

[4] E. Tarkesh Esfahani, *Investigation of Brain Computer Interface as a New Modality in Computer Aided Design/Engineering Systems*, Universidad de California, Riverside. Septiembre 2012.

[5] *Assistive Context-Aware Toolkit (ACAT)*, accesible en: <https://01.org/acat/downloads>

[6] Sitio de descarga del software Plaphoons, <http://plaphoons.softonic.com/>

[7] Sitio de descarga del software de HADASoft, <http://plaphoons.softonic.com/>

[8] Abdulkader S., Ayman Atia, Mostafa-Sami M. Mostafa, *Brain computer interfacing: Applications and challenges*

[9] EMOTIV eStore, application <https://emotiv.com/store/app/>

BIOESTADISTICA APLICADA: SOFTWARE DL50

Marta Ofelia Rivero¹, Acosta Nelson³, Kornuta Carlos Antonio^{2,3,4}

¹ Cátedra de Bioestadística. Dpto. Matemática, Facultad de Ciencias Exactas Químicas y Naturales. Universidad Nacional de Misiones; Posadas

²Departamento de Informática, Facultad de Ciencias Exactas Químicas y Naturales. Universidad Nacional de Misiones; Posadas.

³Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires, Tandil.

⁴Becario Doctoral CONICET.

morivero@fceqyn.unam.edu.ar, nacosta@exa.unicen.edu.ar ckornuta@exa.unicen.edu.ar

Resumen

Este trabajo presenta un software para el estudio de la supervivencia a través del método probit; el cual es utilizado en diversos trabajos de investigación y en tesis de grado.

El software fue desarrollado para estimar los valores de la dosis letal 50, es una herramienta que permite en pocos pasos obtener el cálculo de la Dosis Letal 50 (DL50) por el método probit. Este indicador, representa la dosis de una sustancia o radiación, que es mortal para la mitad de los individuos de la prueba. En la fase de programación del mismo trabajaron alumnos de la carrera licenciatura en Sistemas de Información de la FCEQyN de la UNaM.

Palabras clave: Bioestadística, DL50, Aplicación, software

Contexto

Este proyecto se encuentra inserto dentro de los proyectos prioritarios del departamento de matemática, como fortalecimiento inter-facultad, es decir el trabajo cooperativo entre distintas carreras de la misma unidad académica. Así surge este proyecto entre la carrera de Farmacia y la Licenciatura en Sistemas de Información.

Introducción

Una de las problemáticas con la que se enfrentan los investigadores en ciencias de la salud es el manejo de datos estadísticos de las muestras trabajadas en el laboratorio. Estas muestras en general son grandes y la mayoría de los software son complejos para ellos dada la diversidad de pasos u opciones a tener en cuenta.

Desde esta inquietud un grupo de investigadores de la carrera de Farmacia solicitó la colaboración de la cátedra de Bioestadística con el fin de hallar un software lo más sencillo de manejar para las investigaciones de supervivencia.

Por lo que se diseñó un software a medida, que cumpla con los requerimientos de este tipo de usuarios.

Se comenzó por determinar los pasos para la obtención del indicador DL50.

El primer paso es el conteo de las plantas muertas en cada extracto y cada blanco, luego, se corrigen las mortalidades mediante la fórmula de Abbott y paralelamente se utiliza otra corrección que se basa en el porcentaje de supervivencia de los individuos. Las fórmulas de estas correcciones son las siguientes:

$$M = \frac{m_e - m_b}{1 - m_b}$$

(Formula de Abbott) (Eq. 1)

Donde:

M = Mortalidad.

m_e = mortalidad en el extracto.

m_b = mortalidad en el blanco.

$$m_e = \frac{r}{n} \text{ (Eq.2)}$$

$$m_b = \frac{r'}{n} \text{ (Eq.3)}$$

r = plantas muertas en el extracto.

r' = plantas muertas en el blanco.

n = Número de individuos.

Como el número de individuos es constante (10 en este caso), la fórmula se expresa:

$$\%M = \frac{(m_e - m_b)}{(10 - m_b)} \times 100 \text{ (Eq. 4)}$$

La segunda corrección utilizada es la siguiente:

$$S = \frac{(S' - r)}{S'} \text{ (Eq. 5)}$$

Donde:

S = supervivencia

S' = plantas vivas en el blanco

Mortalidad = 1 - Supervivencia (Eq. 6)

Igualando las dos ecuaciones anteriores:

$$\%M = \frac{r}{S'} \times 100 \text{ (Eq. 7)}$$

A continuación, se determina el DL50 mediante la ecuación de la gráfica Ldp por el método de los mínimos cuadrados. Para verificar la linealidad se realiza una prueba mediante el estadístico de prueba t de Student bilateral (de dos colas) y n-2 grados de libertad, utilizando el valor del factor de correlación lineal (rc), donde la hipótesis nula es que no existe correlación entre Probit y el Log10 dosis. Para ello se utiliza la siguiente ecuación:

$$t = \frac{\sqrt{r_c(n-2)}}{\sqrt{1-r_c^2}} \text{ (Eq. 8)}$$

Para probar la adecuación de la gráfica se realiza el test de bondad de ajuste mediante el estadístico Ji - cuadrado (c 2), en este caso la hipótesis nula es que la línea Ldp es un modelo que se ajusta a los datos.

Para el cálculo del c^2 se utiliza la mortalidad esperada de las larvas (P), la cual se halla con los Probit esperados, calculados a partir de la adecuación de la gráfica Ldp. La ecuación es la siguiente:

$$c^2 = S \frac{(r - n.P)^2}{n.P.(1 - P)}$$

(Eq. 9)

El último paso del método gráfico consiste en establecer los límites de confianza para la DL50, para ello se utilizan los coeficientes de ponderación (W) para cada Probit, que se encuentran tabulados. Las ecuaciones necesarias son las siguientes:

$$S_{\log DL50} = \frac{1}{B} \sqrt{\frac{1}{\sum^N W} + \frac{(\log DL50 - X_{media})^2}{\sum^N W \cdot (X - X_{media})^2}}$$

(Eq. 10)

Donde:

$$X_{media} = \frac{\sum^N W \cdot X}{\sum^N W}$$

(Eq. 11)

$$\begin{aligned} \sum^N W \cdot (x - x_{media})^2 \\ &= \sum^N W \cdot X^2 \\ &= \frac{(\sum^N W \cdot X)^2}{\sum^N W} \end{aligned}$$

(Eq. 12)

El intervalo de confianza es entonces representado como:

$$\log DL50 \pm 1,96 \cdot S_{\log DL50}$$

(Eq. 13)

El valor de 1,96 corresponde a un nivel del 95% de confianza para una distribución normal. Para realizar

estos cálculos se elaboraron tablas en Infostat.

Una vez realizado el método gráfico, se procede a emplear el método de la máxima verosimilitud, que consiste en la suposición de normalidad de la población estudiada, así, se pueden plantear una serie de funciones que denominan los valores de P (Probit) y W (coeficiente de ponderación), con los cuales se pueden hallar los estimativos más próximos para a y b (intercepto en las ordenadas y pendiente), denominados A y B respectivamente, en una línea recta.

1. Con el efecto obtenido, mortalidad en este caso, se encuentra el respectivo valor de Probit (denominado Probit observado o empírico).

2. Se realiza la gráfica de los Probits observados y las dosis metamétricas (X), es decir el Log10 de las dosis.

3. Usando esta línea recta o su ecuación se encuentran los Probits esperados para cada valor de X.

4. Se obtienen los Probits de "trabajo", utilizando para ello las tablas de Probit de trabajo para cada valor de Probit esperado y porcentaje de efecto (% mortalidad).

5. Para cada valor de Probit de trabajo se encuentra el correspondiente valor del coeficiente de ponderación (W).

6. Con estos valores es posible encontrar los estimativos de a y b, correspondientes a los valores A y B, empleando el método de los mínimos cuadrados:

$$\begin{aligned} B &= \frac{\sum W(X - X_{media}) \cdot (Y - Y_{media})}{\sum^N W \cdot (X - X_{media})^2} \\ &= \frac{S_{xy}}{S_{xx}} \end{aligned}$$

(Eq. 14)

$$A = Y_{media} - B_{xmedia}$$

(Eq. 15)

Donde:

$$X_{media} = \frac{\sum^N W.X}{\sum^N .W}$$

(Eq.16)

$$Y_{media} = \frac{\sum^N W.Y}{\sum^N .W}$$

(Eq. 17)

$$S_{xx} = \sum^N W.X^2 - \frac{(\sum^N W.X)^2}{\sum^N W}$$

(Eq. 18)

$$S_{yy} = \sum^N W.Y^2 - \frac{(\sum^N W.Y)^2}{\sum^N W}$$

(Eq. 19)

$$S_{xy} = \sum^N W.XY^2 - \frac{(\sum^N W.X).(\sum^N W.Y)}{\sum^N W}$$

(Eq. 20)

7. Con estos valores para cada X se puede encontrar una segunda aproximación a la línea de regresión.

8. Utilizando esta línea se puede encontrar un segundo grupo de Probitsesperados para cada X y se repite el ciclo desde el paso 3, este procedimiento se repite hasta que no ocurra cambio significativo en la línea. Se puede comprobar la adecuación de la línea mediante el estadístico de prueba c 2.

$$X^2 = S_{yy} - \frac{(S_{xy})^2}{S_{xx}}$$

(Eq. 21)

9. Para la línea final se obtiene el estimativo de la DL50, resolviendo para $Y = 5$, $\text{Log}_{10}DL50 = (5 - A)/B$, o simplemente interpolando en la gráfica lineal final para $Y = 5$. Para hallar la SLogDL50, utilizada para expresar el intervalo de confianza, se utiliza la ecuación:

$$S \log DL50 = \frac{1}{B} \sqrt{\frac{1}{\sum^N W} + \frac{(\log DL50 - X_{media})^2}{S_{xx}}}$$

(Eq. 22)

El intervalo de confianza se expresa igual que en el método gráfico (Eq. 13).

Una vez organizadas las formulas y diseñado los algoritmos, correspondientes a la fase de “análisis del sistema”, del ciclo de vida de un producto software, se procedió a la programación o implementación.

Líneas de Investigación, Desarrollo e Innovación

Dentro de este proyecto se integran las siguientes líneas de Investigación y Desarrollo:

- Bioestadística
- Matemática computacional
- Computación
- Análisis de sistemas

En este sentido, nuestro trabajo tiene como eje resolver a través del empleo de un software de manejo sencillo los problemas que se les presentan a los investigadores del área de ciencias de la salud para el manejo de datos de laboratorio.

Resultados y Objetivos

Este software es utilizado en trabajos de investigación de supervivencia que se basan en el método probit con dosis letal 50.

En las tesinas el software es utilizado para estimar los valores de DL_{50} en las pruebas de toxicidad aguda con múltiples concentraciones por el método Probit (método paramétrico).

El Objetivo General, consiste en evaluar la toxicidad a partir del bioensayo homologado y la determinación de la dosis letal 50 (DL_{50}) y como objetivos específicos, se analizan los resultados obtenidos del bioensayo a partir del programa DL_{50} por el método Probit; además, este software se utiliza para determinar muchos tipos específicos de dosis-respuesta o experimentos de tipo binomial, en una diversidad de campos.

Una vez que se ejecuta una regresión, el investigador puede usar la salida del análisis probit con la finalidad de comparar la cantidad de producto químico necesario para crear la misma respuesta en cada uno de los diversos productos químicos. Existen muchos criterios de valoración utilizados para comparar la toxicidad de las sustancias químicas diferentes, pero el LC_{50} (líquidos) o DL_{50} (sólidos) son los resultados más ampliamente utilizados de los modernos experimentos dosis-respuesta. Este software nos permite a partir de la concentración de las dosis suministradas obtener en simples pasos el valor del probit.

Esta herramienta, además de realizar los cálculos necesarios, a partir del ingreso de los datos de la concentración de la dosis administrada y el número de sujetos muertos, obtiene determinados valores estadísticos como ser: la media, la

desviación estándar, los límites de confianza y el probit.

En este contexto, el software se sigue utilizando en tareas de investigación sobre la Lesmaniasis y en las tesis de grado de los alumnos de la carrera Farmacia, correspondiente a la Facultad de Ciencias Exactas, Químicas y Naturales de la Universidad Nacional de Misiones.

Formación de Recursos Humanos

El equipo de trabajo está conformado por un director (categoría III, del programa de incentivos SPU), profesor de bioestadística, un becario Doctoral CONICET de la Universidad del Centro de la Provincia de Buenos Aires, doctorando en Matemática Computacional e Industrial, dos auxiliares informáticos, y alumnos de la carrera licenciatura en sistemas de Información y dos tesistas de grado de la carrera en Farmacia. Las tesis han sido presentadas y aprobadas. Una de ellas versa sobre la determinación de la dosis letal 50 del Glifosato bajo el modelo de ensayo de toxicidad aguda de *Lactuca Sativa* y la otra sobre el Cálculo de la DL_{50} de deltametrina, aplicación de protocolo de bioensayo agudo de letalidad sobre *Artemia Salina*.

Bibliografía

- Finney, D.J, (1952) Probit Analysis. Cambridge University Press
- Greenberg, B.G (1980) Chester Bliss, 1899-1979, International Statistical Review/ revue Internationale de statistique 8 (1) 135-136.
- Hahn, E.D and R Soyer (2008). Probit and logit models: Differences in a multivariate Realm.

- Arregui, María Cristina y otros. (2010). Informe a cerca del grado de toxicidad del glifosato. Universidad Nacional del Litoral, Santa Fé. Capítulo 2, p.17-27. Disponible en: <http://www.unl.edu.ar/noticias/media/docs/Informe%20Glifosato%20UNL.pdf>. Capturado el: 10/10/12.
- Repetto Jiménez, Manuel; Repetto Kuhn, Guillermo. (2009). Toxicología Fundamental. Cuarta edición. Editorial Diaz de Santos. Capitulo 2, p. 23-32

Living Labs en la región Noroeste de la provincia de Buenos Aires

H. Ramón, S. Pompei, F. López Gil, P. Occhipinti, D. Perez

Instituto de Investigación y Transferencia de Tecnología (ITT),
Escuela de Tecnología
Universidad Nacional del Noroeste de la Provincia de Buenos Aires
Junín (B)

{hugo.ramon, sabrinapomei, fernandolopezgil, pocchipinti, danielaperez}@itt.unnoba.edu.ar

Resumen

A fines de los años 90 surge el concepto de living labs como una metodología de investigación centrada en el usuario. Este concepto fue adoptado por distintas comunidades científicas las cuales reinventaron el concepto de laboratorio viviente o living lab definido hoy como un ecosistema de innovación abierta basado en la co-creación integrando procesos de investigación e innovación en contextos de la vida real.

A diferencia de los típicos de los modelos de innovación denominados de Triple Hélice, las redes de living labs introducen a un nuevo actor, los ciudadanos y forman una Cuarta Hélice.

Al ser esta una propuesta de innovación e investigación relativamente nueva se pretende con la presente línea determinar el estado del arte focalizando en la estructura adoptada por los laboratorios vivientes, metodologías empleadas, formas de gestión y operación para converger en el proceso asociado.

Todo esto nos permitirá luego desarrollar o adaptar un modelo de referencia para la instalación y desarrollo de un laboratorio viviente, el cual se pretende validar en la región del Noroeste de la Provincia de Buenos Aires.

Palabras clave:

Laboratorio viviente, innovación abierta, Procesos

Contexto

La presente línea está inserta en el proyecto de investigación "Certificación de Calidad y Digitalización de Procesos en Organizaciones Tecnológicas de la región UNNOBA", acreditado y financiado por la Universidad Nacional del Noroeste de Buenos Aires (UNNOBA) en la convocatoria "Subsidios de investigación bianuales" 2015, con lugar de trabajo en el Instituto de Tecnología y Transferencia (ITT).

Introducción

A fines de los años 90 William J. Mitchell, profesor de arquitectura del

Massachusetts Institute of Technology (MIT), desarrolló el concepto de Living Labs como una metodología de investigación centrada en el usuario para probar, validar, realizar prototipos y refinar soluciones complejas en entornos reales. Mitchell estaba particularmente interesado en involucrar activamente a los ciudadanos en la planificación y el diseño urbano.

La idea de incluir al usuario en los procesos de innovación y ubicarlo en una posición centralizada están dando lugar a la aparición de nuevos espacios de interacción y comunicación en los que el consumidor deja de considerarse en forma pasiva para convertirse en un actor activo [1].

Esta idea de involucrar al ciudadano en el proceso de diseño fue tomada y adaptada luego por distintas comunidades europeas de investigadores reinventando así el concepto de Living Labs [2].

Por su parte, la Red Europea de Living Labs (ENoLL) define Living Labs como ecosistemas de innovación abierta basados en la co-creación integrando procesos de investigación e innovación en contextos de la vida real [3].

De acuerdo los Living Labs surgen tanto en Estados Unidos como también en el norte de Europa promovidos por centros de investigación y grandes empresas, todos ellos pertenecientes al sector TIC. Posteriormente, impulsados por la ENoLL, se van difundiendo por el resto de Europa y otros continentes. En este proceso surge la idea de laboratorio

ciudadano, en cierta medida como sinónimo de Living Labs, aunque poniendo acento en el entorno urbano, involucrando nuevos actores no necesariamente del sector TIC.

Según [4] *“En América Latina el proceso de surgimiento de los laboratorios ciudadanos fue más tardío y presenta otros matices, dado que, entre otros factores, la mayor parte de los mismos no han sido promovidos por destacados centros de investigación, ni por grandes empresas pertenecientes al sector TIC.”* Y en las reflexiones finales indica, *“...ninguna de las experiencias identificadas en Argentina surge con el propósito de crear un laboratorio ciudadano y su análisis permite constatar que no todas ellas trabajan con una metodología de innovación abierta; aunque sí lo hacen, en mayor o menor medida, centrando su accionar en las necesidades y requerimientos de los usuarios ..”*.

En el caso de Argentina, en noviembre del año 2011, a fin de instalar el debate sobre la problemática Living Labs y comenzar a reconocer experiencias en curso en el país, se realizó la Primer Jornada Argentina sobre Living Labs organizada por el centro REDES (Centro de Estudios sobre Ciencia, Desarrollo y Educación Superior).

A diferencia de los típicos de los modelos de innovación de los modelos de innovación denominados de Triple Hélice [5], las redes de living labs introducen a un nuevo actor, los ciudadanos y el ambiente [6] [7] [8].

De acuerdo [9], se concluye que el movimiento de Living Labs en términos de teoría e investigación empezó en 2006, pero en términos de impacto y calidad, todavía es insignificante. Hay estudios comparativos que se centran en el valor añadido de Living Labs y los pone como un campo de investigación en desarrollo que requiere más investigación con el fin de avanzar en la generación de conocimiento.

En el apartado de Results & discussion del mismo trabajo resaltan:

“When going through all the papers, two important issues arise. First, only a small minority of the papers reports on well-grounded empirical research on Living Labs. The majority of the papers are descriptive single or multiple case studies, or conceptual papers relying on desk research, without a rigid methodology being used or explained.”

Este contexto hace necesario el estudio de los procesos existentes que soportan la metodología de Living Labs analizando las modificaciones culturales regionales necesarias, relacionando estas con los indicadores de innovación muy importantes en especial en la Región UNNOBA [10] [11][12][13].

Dada la necesidad de acercar la ciencia a la sociedad como así también la de estimular la participación ciudadana en los procesos de innovación y desarrollo local, y considerando la situación previamente descripta surgen los siguientes cuestionamientos:

- ¿Cuáles son las razones que provocan el tardío surgimiento de Living Labs en Argentina?
- ¿Es posible crear un laboratorio ciudadano con este propósito inicial en nuestro país?
- ¿Cuáles son los requerimientos necesarios para la creación e implementación de un living lab para la región del noroeste de la provincia de Buenos Aires?
- ¿Cómo afectan estos laboratorios en la normativa de patentes?
- ¿Existen procesos que puedan adaptarse a la cultura regional?

Líneas de Investigación, Desarrollo e Innovación

El tema de nuestro trabajo de investigación sobre living labs, ecosistemas de innovación abierta basados en la co-creación integrando procesos de investigación e innovación en contextos de la vida real [14][15][16].

Dentro de este tema se trabajarán los siguientes ejes:

- Estrategias para el fomento de la participación ciudadana y la creación de relaciones sinérgicas entre comunidad, laboratorio, gobierno y sector privado.
- Enfoque y metodologías de investigación de los Living Labs.
- Gestión, operación y sostenibilidad de los Living Labs.

Resultados y Objetivos

El objetivo principal de este proyecto es definir un modelo de referencia que promueva la creación e implementación de un laboratorio viviente. Para ello se propone:

- Elaborar el estado del arte
- Desarrollar/Adaptar el modelo de referencia
- Validar la propuesta en la región del Noroeste de la Provincia de Buenos Aires
- Analizar resultados

Formación de Recursos Humanos

El equipo de trabajo de la presente línea está compuesto por 4 docentes investigadores del área de Sistemas y 3 estudiantes avanzados de la misma área, junto con otros actores de la Región Noroeste de la pcia de Bs. As. cómo son Municipios, Polos Tecnológicos y empresas privadas.

Actualmente se cuenta con una propuesta de tesis de grado en curso para la Lic. en Sistemas de información (UNNOBA), una propuesta de tesis de posgrado para la Maestría en Metodologías de la Investigación Científica (UNLa) y una tesis Especialización de la Innovación y la Vinculación Tecnológica (UNNOBA).

Referencias

[1] Usuarios activos, living labs e innovación abierta. El caso del citilab de Cornellà. Dr. Jordi Colobrans. X

CONGRESO ESPAÑOL DE SOCIOLOGÍA. Grupo de trabajo: 19 - Sociología del Consumo. Pamplona 1-3 julio 2010.

[2] Jonathan Wallace, Maurice Mulvenna, Brendan Galbraith, Suzanne Martin (TRAIL), Francesco Molinari (RT), Tiina Ferm (TUAS), Jesse Marsh (RS), John Heaven (TUTECH), Sven Besteher (HAMBURG), Iolanda Romano, Emanuela Galetto (AU); “Living Lab Methods and Tools for the EU Public Sector”. Electronic Participation Tools for Spatial Planning and Territorial Development (PARTERRE), dic-10, http://www.parterre-project.eu/file.php/?fname=PARTERRE_D2.3.pdf&codice=5389dda6fc_f12

[3] Red Europea de Living Labs (ENoLL), <http://www.openlivinglabs.eu/aboutus>, enero 2016.

[4] Schiavo, Ester; dos Santos Nogueira, Camilla; Vera, Paula; “Entre la divulgación de la cultura digital y el surgimiento de los laboratorios ciudadanos. El caso argentino en el contexto latinoamericano”. Revista Iberoamericana de Ciencia, Tecnología y Sociedad - CTS, vol. 8, núm. 23, mayo, 2013, pp. 179-199, Centro de Estudios sobre Ciencia, Desarrollo y Educación Superior, Buenos Aires, Argentina

[5] The dynamics of innovation: from National Systems and ‘Mode 2’ to a Triple Helix of university–industry–government relations. Etzkowitz, H. & Leydesdorff, L. (2000) Research Policy.

[Online] 29 (2), 109–123. Available from: doi:10.1016/S0048-7333(99)00055-4

[6] ‘Mode 3’ and ‘Quadruple Helix’: toward a 21st century fractal innovation ecosystem. Carayannis, E.G. & Campbell, D.F. (2009) International Journal of Technology Management. 46 (3), 201–234.

[7] Exploring Quadruple Helix. Outlining user-oriented innovation models, University of Tampere, Work Research Center, Working Paper N° 85. ARNKIL, R., 2010. www.cliqproject.eu/

[8] Facilitating Quintuple Helix innovation with Urban Living Labs. Bastiaan Baccarne, Dimitri Schuurman, Lieven De Marez. iMinds – MICT – Ghent University Korte Meer Ghent. Belgium. XXVI ISPIM Conference – Shaping the Frontiers of Innovation Management, Budapest, Hungary on 14-17 June 2015. www.ispim.org.

[9] Living Labs: a systematic literature review. Dimitri Schuurman, Lieven De Marez & Pieter Ballon. , ISBN (e-book): 9789082102741, 2015 ENoLL - European Network of Living Labs.

[10] State-of-the-art in utilizing Living Labs approach to user-centric ICT innovation 13 (13), Mats Eriksson, CDT, Luleå University of Technology, Sweden http://www.vinnova.se/upload/dokument/verksamhet/tita/stateoftheart_livinglabs_riksson2005.pdf

[11] Futures of Innovation Systems and Systemic Innovation Systems: Towards

Better Innovation Quality with New Innovation Management Tools. Writer & Finland Futures Research Centre, University of Turku, ISBN 978-952-249-126-8, ISSN 1797-132, Jari Kaivo-oja http://www.utu.fi/fi/yksikot/ffrc/julkaisut/e-tutu/Documents/eTutu_2011_8.pdf

[12] Enhancing Co-Creation with Privacy and Security-by-Design methodologies , Francesco Alberti, Sauro Vicini, ISBN (e-book): 9789082102741, 2015 ENoLL - European Network of Living Labs.

[13] The Measurement of Scientific and Technological Activities. Proposed Guidelines for Collecting and Interpreting Technological Innovation Data. Oslo Manual, 3rd edition, 2005, ISBN 978-92-64-01308-3 www.oecd.org/sti/oslomanual

[14] Almirall, Esteve; Lee, Melissa y Wareham, Jonathan; “Mapping Living Labs in the Landscape of Innovation Methodologies”. Technoogy Innovation Management Review, Septiembre de 2012, (12-18).

[15] Anna Ståhlbröst and Marita Holst, “The Living Lab methodologyc handbook”. Social Informatics at Luleå University of Technology and CDT – Centre for Distance-spanning Technology, Sweden, 2012.

[16] Anna Ståhlbröst and Marita Holst, “The Living Lab methodologyc handbook”. Social Informatics at Luleå University of Technology and CDT – Centre for Distance-spanning Technology, Sweden, 2012.

Performance de arquitecturas multiprocesador: técnicas de modelado y simulación y plataformas reconfigurables.

Diego Encinas^{1,2}, Eduardo Kunysz¹, Jimena Jara¹, Martin Morales^{1,3}

¹Instituto de Ingeniería y Agronomía - Universidad Nacional Arturo Jauretche

²Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática
- UNLP

³Laboratorio de Ingeniería en Sistemas de Información - FRLP - UTN

dencinas@unaj.edu.ar, ekunysz@unaj.edu.ar, elchejime@gmail.com,
martin.morales@unaj.edu.ar

Resumen

El objetivo de esta línea de investigación es el estudio de la performance de las arquitecturas multiprocesador a través de modelos de simulación y plataformas reconfigurables. Enfocando a la obtención de herramientas que permitan predecir la eficiencia del sistema ante posibles escenarios y reconfigurar el sistema en tiempo real. Analizando los diferentes componentes del sistema que pueden influir en las prestaciones significativamente y pueden llegar a modelarse y/o reconfigurarse.

Palabras clave: *Arquitecturas Multiprocesador. Simulación. Sistema E/S paralela. Modelado y Simulación basado en agentes (Agent-Based Modeling and Simulation, ABMS). Plataformas reconfigurables para procesamiento paralelo. Lógica programable.*

Contexto

Se presenta una línea de Investigación que es parte del Proyecto de Investigación “Modelado y Simulación en Cómputo de Altas Prestaciones (HPC). Aplicaciones en arquitecturas multiprocesador, sistemas paralelos y redes de datos” de la Universidad Nacional Arturo Jauretche (UNAJ), acreditado por resolución interna 186/15.

En el tema existe un convenio de colaboración en actividades de Investigación y Postgrado con el Instituto de Investigación en Informática – LIDI y el Centro de Técnicas Analógico-Digitales – CeTAD de la Universidad Nacional de La Plata.

Por otra parte, se tiene financiamiento en el marco del programa “Universidad, Diseño y Desarrollo Productivo” del Ministerio de Educación a través de varios proyectos aprobados en la UNAJ.

Introducción

El crecimiento sostenido en la demanda del poder de cómputo remarca

la necesidad de sistemas con enfoques de paralelización masiva y cómputo de alta performance (HPC, High Performance Computing) [1]. Los clusters se han convertido en uno de los enfoques principales para lograr paralelismo a bajo costo. Una noción extendida lo constituye la utilización de grid computing y más recientemente cloud computing. Independientemente de la solución, estos sistemas constan de un gran número de componentes incluyendo nodos de procesamiento, bancos de memoria, discos, entre otros.

Sistemas de E/S Paralela

Las exigencias en los sistemas de E/S paralelos se han incrementado debido al aumento en número, velocidad y potencia de las unidades de procesamiento en los clusters. También las aplicaciones científicas que utilizan cómputo de altas prestaciones acrecientan estos requerimientos.

En muchos casos, el cuello de botella de los sistemas paralelos es la E/S de estos sistemas dada las exigencias que debe afrontar [2]. La E/S Paralela es esencial para emparejar el avance de las arquitecturas de los procesadores y el rápido crecimiento de la capacidad computacional. Aunque la arquitectura jerárquica de memoria multinivel puede evitar grandes pérdidas de prestaciones debido a los retardos de acceso a disco, la capacidad de memoria es limitada. Además, como la capacidad computacional aumentará, la disponibilidad de memoria por core decrecerá, especialmente si la escala de los sistemas de HPC se proyecta a millones de cores o más. Varias simulaciones científicas y de ingeniería de áreas críticas de investigación, tales como la nanotecnología, astrofísica, clima y energía física están convirtiéndose en

aplicaciones intensivas de datos. Para poder disminuir la brecha entre CPUs-E/S se deben identificar los factores que influyen en las prestaciones y proponer nuevas soluciones [3] [4].

En el área de tolerancia a fallas en sistemas de cómputo de alta prestaciones se puede notar la importancia de la unidad de E/S en las arquitecturas paralelas como un punto a mejorar para lograr cubrir las exigencias de las aplicaciones que utilizan HPC. Una manera de llevar a cabo este trabajo es utilizar técnicas de simulación para evaluar el efecto de los cambios de los factores con mayores influencias en las prestaciones del sistema de E/S paralelo.

Se puede disminuir la complejidad y la probabilidad de errores en la generación de sistemas híbridos desarrollando una simulación específica de éstos utilizando diferentes frameworks [5] [6] [7].

Las aplicaciones científicas con un uso intensivo de datos utilizan software de E/S paralelo para acceder a archivos. Contar con una herramienta que permita predecir el comportamiento de este tipo de aplicaciones en HPC es de gran utilidad para los desarrolladores de aplicaciones paralelas. Por otro lado, ABMS ha sido utilizado para modelar problemas y sistemas complejos en diversas áreas de la ciencia.

Evaluar las prestaciones del subsistema de E/S con diferentes configuraciones y la misma aplicación permite adaptar la configuración de E/S teniendo en cuenta el patrón de acceso de la aplicación Pero también puede ser una gran ventaja analizar las necesidades de las aplicaciones antes de configurar el sistema físico. Una manera de predecir el comportamiento de las aplicaciones en el sistema de cómputo ante distintas configuraciones, es utilizando técnicas de modelado y simulación.

Se propone modelar y simular la arquitectura de E/S paralela, por medio de técnicas de simulación basadas en agentes o Sistemas Multi-Agente, (MAS-Multi Agent Systems), para evaluar el efecto de dimensionar el sistema de E/S o cambiar componentes como la red de almacenamiento, dispositivos de E/S, entre otros [8].

Plataformas reconfigurables para procesamiento paralelo

Con el avance en la búsqueda de mayores prestaciones, la tecnología de procesamiento ha evolucionado hacia la utilización de múltiples procesadores. No solo la utilización de “multi-cores” en una oblea de silicio, sino la utilización de procesadores de aplicaciones específicas como placas de video (GPUs) o co-procesadores matemáticos agregados al sistema por buses de alta velocidad como PCI. El desafío en estos casos por ejemplo, consiste en trasladar el problema concreto a las limitaciones de la programación gráfica.

En los últimos años, se ha buscado expandir este concepto hacia plataformas de procesamiento más específico. Para obtener mayor eficiencia, los fabricantes de computadoras de altas prestaciones, han introducido unidades de co-procesamiento de arreglos de lógica programable. La aplicación principal se ejecuta en los microprocesadores, mientras que las FPGAs manejan las porciones de código que requieren mayor tiempo de ejecución [9] [10].

Una de las ventajas de este tipo de arquitectura es la capacidad que tiene el procesador principal de reconfigurar las FPGAs en tiempo real. Esto permite la reutilización de partes de hardware en virtud de mejorar la performance del software, lo que crea un nuevo paradigma

en el campo de los desarrolladores de aplicaciones

Típicamente el clock de una FPGA es un orden de magnitud o más lento que el de un procesador. Sin embargo desde el punto de vista de la velocidad, las FPGAs obtienen su ventaja en los siguientes tres factores:

- *Intensidad*: en el mejor de los casos un CPU puede realizar una operación entre enteros cada dos ciclos de reloj (caso ideal en donde se estaría utilizando cache, y con el pipeline funcionando sin interrupciones). El peor caso es significativamente más desfavorable. En una FPGA se implementa sólo la funcionalidad necesaria para la aplicación particular. Por lo tanto se prescinde del “overhead” producto de la arquitectura sofisticada que tiene un CPU de propósitos generales.
- *Latencia*: con una FPGA se puede obtener una granularidad fina sobre el control de donde se encuentran los datos en memoria.
- *Paralelismo espacial*: se puede realizar un pequeño pipeline de propósitos especiales para realizar una operación particular, y luego replicarlo dentro del chip de la FPGA [11].

Se propone modelar y simular estas características que demuestran una mejora considerable en el procesamiento de cálculo intensivo aplicado a clusters [12].

Líneas de Investigación, Desarrollo e Innovación

Temas de Estudio e Investigación

- Arquitecturas multiprocesador para procesamiento paralelo: multiprocesador de memoria compartida, multiprocesador on-chip de memoria distribuida. Multicore, Clusters, Clusters de multicore. Grid. Cloud.
- Arquitectura de E/S paralela considerando el software, hardware, comunicaciones entre módulos y dispositivos de almacenamiento.
- Modelado y simulación basada en agentes.
- Técnicas de comunicación entre dispositivos de lógica programable para optimizar el rendimiento y consumo de las HPRC (High Performance Reconfigurable Computing).
- Generar múltiples dispositivos reconfigurables para implementar sistemas paralelos con múltiples unidades de procesamiento.
- Simular distintas topologías de interconexión entre unidades de procesamiento reconfigurables para optimizar el camino de la información.

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D se participa en el dictado de la carrera de Ingeniería Informática de la UNAJ. También aportan trabajos de alumnos de las materias Redes de Computadoras 2, Programación en Tiempo Real, Sistemas Operativos 2 y Sistemas Distribuidos.

En esta línea de I/D existe cooperación a nivel nacional e internacional. Hay 3 investigadores realizando su Doctorado y 3 alumnos avanzados de grado colaborando en las tareas.

Resultados y Objetivos

Investigación experimental

- Diseño y desarrollo de modelos mediante técnicas de Modelado y simulación basada en agentes (ABMS)
- Utilización de agentes para generar la funcionalidad de los elementos físicos (procesadores, memoria, buses, drivers, entre otros) como así también de las interfaces en las arquitecturas de E/S.
- Análisis y modelado de librerías de archivos para aplicaciones que utilizan computo de altas prestaciones-HPC.
- Desarrollo de pruebas de conceptos con el modelo inicial del sistema de archivos paralelos utilizando un entorno de programación MAS.

Referencias

1. Grama A, Gupta A, Karypis G, Kumar V. "Introduction to parallel computing". Second Edition. Pearson Addison Wesley, 2003.
2. H Hennessy, J. L., Patterson, and D. A., Computer Architecture, Fourth Edition: A Quantitative Approach. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2006.
3. J. M. May, Parallel I/O for high performance computing. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001.
4. V. Balaji, Earth system modelling – Volume 4. IO and Postprocessing. Springer, 2013.
5. D. Encinas, Utilización de un reloj global para el modelado de un ambiente simulado

- distribuido. XVIII Congreso Argentino de Ciencias de la Computación. 2012
6. D. Encinas, Simulación de una red CAN para dimensionar las comunicaciones de una IMU. VII Congreso Argentino de Tecnología Espacial. 2013.
 7. D. Black, SystemC: From the Ground Up. Second Edition, Springer, 2010.
 8. D. Encinas et al., Modeling I/O System in HPC: An ABMS Approach. The Seventh International Conference on Advances in System Simulation (SIMUL), ISBN: 978-1-61208-442-8, 2015
 9. K. Underwood “FPGAs vs. CPUs: Trends in Peak Floating-Point Performance”, Sandia National Laboratories, 2004 .
 10. C. Chang, J. Wawrzynek, and R. W. Brodersen, “BEE2: A High-End Reconfigurable Computing System”, IEEE, University of California, Berkeley, 2005.
 11. E. J. Kunysz, J. A. Rapallini, J. Osio, “Sistema de cómputo reconfigurable de alta performance (Proyecto HPRC) , 3ras Jornadas ITE -Facultad de Ingeniería – UNLP, 2015.
 12. K. Kintali and Y. Gu, "Model-Based Design with Simulink, HDL Coder, and Xilinx System Generator for DSP", MathWorks, White Paper (<http://www.mathworks.com/fpga-design/simulink-with-xilinx-system-generator-for-dsp.html>), 2012.

Performance de cloud computing para HPC: despliegue y simulación.

Brian Galarza¹, Gonzalo Zaccardi¹, Daniel Rossatto¹, Román Bond¹, Martin Morales^{1,2}, Diego Encinas^{1,3}

¹Instituto de Ingeniería y Agronomía - Universidad Nacional Arturo Jauretche

²Laboratorio de Ingeniería en Sistemas de Información - FRLP - UTN

³Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP

gezaccardii@gmail.com, briangalarza@hotmail.com, danielrosatto@gmail.com, roman.alejandro.b@gmail.com, martin.morales@unaj.edu.ar, dencinas@unaj.edu.ar

Resumen

El objetivo de esta línea de investigación es el estudio de la performance de las arquitecturas tipo cloud a través del despliegue de IaaS y simulación para la ejecución de aplicaciones, en particular en el área de cómputo paralelo de altas prestaciones (HPC). Enfocando a la obtención de herramientas que permitan predecir la eficiencia del sistema ante posibles escenarios. Analizando los diferentes componentes del sistema que pueden influir en las prestaciones significativamente y pueden llegar a modelarse y/o configurarse.

Palabras clave: *Arquitecturas Multiprocesador. Cloud Computing. OpenStack. Simulación. CloudSim.*

Contexto

Se presenta una línea de Investigación que es parte del Proyecto de Investigación “Modelado y Simulación en Cómputo de

Altas Prestaciones (HPC). Aplicaciones en arquitecturas multiprocesador, sistemas paralelos y redes de datos” de la Universidad Nacional Arturo Jauretche (UNAJ), acreditado por resolución interna 186/15.

Por otra parte, se tiene financiamiento en el marco del programa “Universidad, Diseño y Desarrollo Productivo” del Ministerio de Educación a través de varios proyectos aprobados en la UNAJ.

Introducción

Cloud Computing es un paradigma que está en constante crecimiento durante estos últimos años, cada vez más compañías y grupos de investigación trabajan en conjunto con el fin de explotar las oportunidades ofrecidas por el mismo [1]. Dicho paradigma ofrece muchas ventajas, tales como el bajo costo de implementación, ya que no se necesitan computadoras de última tecnología debido a que éstas trabajan conjuntamente (Clustering) con la posibilidad de escalar

horizontalmente de manera sencilla. Además, hay software Open Source disponible para los nodos en el clúster como las infraestructuras Eucalyptus, OpenNebula, CloudStack u OpenStack integradas con GNU/Linux y compatibles, por ejemplo, con Amazon WebServices.

En cuanto a las herramientas de simulación, CloudSim es un framework desarrollado en Java que provee las APIs necesarias para que el usuario genere una simulación de un Data Center funcionando como servidor de nube capaz de simular la ejecución de CloudLets. WorkflowSim es un desarrollo basado en CloudSim al que se le ha integrado la ejecución de Workflows. También se utilizará el benchmark IOR para medición de performance.

Análisis de performance en arquitecturas de Cloud Computing

Las comunicaciones en Cloud Computing son una parte fundamental del paradigma que consisten en utilizar distintos nodos y lograr hacerlos funcionar conjuntamente.

Para lograr una comunicación entre estos nodos se propone utilizar OpenStack [2].

OpenStack es un software de código abierto que permite la implementación de, por ejemplo, una “Infrastructure as a Service” (IaaS) a través de múltiples servicios que, de manera coordinada, cumplen diferentes propósitos para lograr el correcto funcionamiento de dicha infraestructura. Algunos de los servicios ofrecidos por OpenStack son: hypervisor (Nova), autenticación (Keystone), Imágenes (Glance), Dashboard (Horizon) y block storage (Cinder). Según las necesidades se pueden requerir de ciertos servicios u otros. La Arquitectura básicamente consiste en dos tipos de

nodos: “Compute Node” y “Controller Node”. Se llaman Compute Node a todos aquellos que se encargan del procesamiento de servicios específicos mientras que Controller Node es aquel que comunica a cada uno de los anteriores [3] [4] [5].

La implementación de estas infraestructuras ofrece ventajas en las cuales los clústers virtualizados trabajan en conjunto ofreciendo un buen rendimiento a bajos costos y con posibilidad de escalabilidad al poder agregar mayor cantidad de nodos para procesamiento de manera sencilla.

Fuel es una herramienta Desarrollada por Mirantis en la cual se ejecuta un script que permite configurar, de manera más amigable respecto a OpenStack, los recursos que se desean otorgar a la infraestructura, desde la cantidad de nodos, los núcleos de procesador, la memoria RAM, entre otros [6].

Fuel trabaja con un nodo master el cual es el encargado de controlar a los nodos slaves que contendrán la infraestructura OpenStack. Es decir, desde el nodo Fuel Master se indican qué paquetes se van a instalar en cada nodo slave (Glance, Nova-Compute, Keystone, etc.) para luego en los slaves tener armados los nodos compute y controller, sin necesidad de realizar configuraciones manuales en cada uno de los mismos.

Simulación de arquitecturas de Cloud Computing

La herramienta para el modelado de la nube Cloudsim [7], se basa en la simulación de todos los componentes de una nube por separado. Proporcionando para cada elemento una clase integrada en la API para que el usuario la integre en la compilación. Las clases o componentes de la nube que provee Cloudsim son las necesarias para soportar cualquier tipo de

Cloud, ya sea SaaS, PaaS o IaaS. El motor de simulación de CloudSim está compuesto por el código de usuario desarrollado y debajo se encuentran las APIs de CloudSim. Dentro del desarrollo del usuario se deben encontrar las especificaciones de la simulación y las políticas de planificación. Para el caso de las APIs de CloudSim se descomponen en Estructuras de interfaces de usuario, servicios de máquinas virtuales, servicios de nube, recursos de la nube y capa de red.

El motor de simulación es capaz de simular el tiempo de ejecución de las apps ingresadas como Cloudlets con información básica.

WorkflowSim [8] un software que se desarrolló utilizando el modelo de sistema de manejo de flujo de trabajo (WMS, en inglés) similar a Pegasus WMS [10] aplicables en sistemas de cálculo en paralelo. Este software provee el motor de simulación de CloudSim y la capa de ejecución de Workflows, necesaria para correr benchmarks del tipo IOR y NAS BT-IO [11] en formato workflow sintético. Estos benchmark tienen como objetivo medir el rendimiento de sistemas de archivos paralelos I/O a nivel POSIX y MPI-IO.

Líneas de Investigación, Desarrollo e Innovación

Temas de Estudio e Investigación

- Arquitecturas multiprocesador para procesamiento paralelo: multiprocesador de memoria compartida, multiprocesador on-chip de memoria distribuida. Multicore, Clusters, Clusters de multicore. Grid. Cloud.
- Plataformas de software para implementar y administrar Clouds públicos, privados e híbridos.
- Simulaciones con CloudSim, WorkflowSim, equivalencias entre los componentes de la arquitectura multiprocesador y la arquitectura simulada.
- Generación de Workflow, utilización WorkflowGenerator [9] para aplicaciones específicas.

Resultados y Objetivos

Investigación experimental

- Implementación de un IaaS encargado de realizar operaciones en procesamiento paralelo aumentando la eficiencia y reduciendo los costes generados.
- Implementación de OpenStack Dashboard y de un sistema propio desarrollado para poder controlar/administrar de manera visual (web) y más básica cada uno de los servicios.
- Utilización de Fuel para administrar OpenStack como sistema de administración de nube (Cloud Computing) a partir de la infraestructura de 2 nodos compute y 1 controller.
- Utilización de CloudSim como herramienta de modelado para la simulación de la arquitectura utilizada pudiendo verificar estadísticas temporales.
- Utilización de WorkflowSim para la ejecución de la simulación con un workflow que pueda ejecutarse en ambos sistemas.

- Implementar un caso de estudio con CloudSim, llamando a un programa en Java que invoque a IOR HPC y ejecutar la simulación con WorkflowSim.

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D se participa en el dictado de la carrera de Ingeniería Informática de la UNAJ. También aportan trabajos de alumnos de las materias Redes de Computadoras 2 y Programación en Tiempo Real.

En esta línea de I/D existe cooperación a nivel nacional. Hay 2 investigadores realizando su Doctorado y 6 alumnos avanzados de grado colaborando en las tareas.

Referencias

1. Kondo, D., Javadi, B., Malecot, P., Cappello, F., Anderson, D. P.: "Cost-benefit analysis of Cloud Computing versus desktop grids". In: IPDPS '09 Proceedings. IEEE International Symposium on Parallel and Distributed Processing. Washington, USA (2009).
2. OpenStack Cloud Software: Open source software for building private and public clouds. <http://www.openstack.org>. Febrero 2015.
3. Galarza, B.; Tuamá, C.; Zaccardi, G.; Encinas, D.; Morales, M. "Implementaciones de Cloud Computing y aplicaciones en el ámbito universitario". I Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2013). Ciudad de Córdoba, Argentina.
4. Zaccardi, G.; Galarza, B.; Encinas, D.; Morales, M. "Implementación de Cloud Computing utilizando OpenStack". II Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2014). Ciudad de San Luis, Argentina.
5. Galarza, B.; Zaccardi, G.; Encinas, D.; Morales, M. "Análisis de despliegue de una IaaS utilizando Openstack". XXI Congreso Argentino de Ciencias de la Computación (CACIC 2015). Ciudad de Junín, Argentina.
6. OpenStack Deployment Fuel. <https://www.mirantis.com/products/mirantis-openstack-software/openstack-deployment-fuel/>. Febrero 2016
7. Rodrigo N. Calheiros , Rajiv Ranjan, Anton Beloglazov, César A. F. De Rose and Rajkumar Buyya "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms" Published online 24 August 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/spe.995.
8. Weiwei Chen , Ewa Deelman - Information Sciences Institute-University of Southern California Marina del Rey, CA, USA - WorkflowSim: A Toolkit for Simulating Scientific Workflows in Distributed Environments.
9. <https://confluence.pegasus.isi.edu/display/pegasus/WorkflowGenerator> To facilitate evaluation of workflow algorithms and systems on a range of workflow sizes, we have developed a set of synthetic workflow generators. Febrero 2015.
10. E. Deelman, et al., Pegasus: Mapping scientific workflows onto the Grid. Lecture Notes in Computer Science: Grid Computing, pp. 11–20, 2004.
11. Parkson Wong, Rob F. Van der Wijngaart - NAS Parallel Benchmarks I/O Version 2.4 - Computer Sciences Corporation NASA Advanced Supercomputing (NAS) Division - NASA Ames Research Center, Moffett Field, CA 94035-1000 - NAS Technical Report NAS-03-002 January 2003.

Evaluación de arquitecturas distribuidas de commodity basadas en software libre

Nelson Rodríguez¹, María Murazzo², Adriana Martín³, Susana Chávez⁴, Diego Medel⁵, Jorge Mercado⁶

¹⁻⁵Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.

⁶Departamento de Matemática - Facultad de Ingeniería - UNSJ

Complejo Islas Malvinas. Cereceto y Meglioli. 5400. Rivadavia. San Juan, 0264 4234129

¹nelson@iinfo.unsj.edu.ar ²marite@unsj-cuim.edu.ar ³arianamartinsj@gmail.com

⁴schavez@iinfo.unsj.edu.ar ⁵vdiego.unsj@hotmail.com ⁶jorgenmp@gmail.com

Resumen

La evolución de las tecnologías asociadas a la conectividad y al procesamiento distribuido ha permitido que surjan nuevas soluciones y necesidades por parte de los usuarios. Nuevos modelos de negocios están emergiendo gracias a la infraestructura de tecnologías informáticas distribuidas como Cluster y Cloud. La variedad de arquitecturas distribuidas ofrecen distintas prestaciones para centros de datos pequeños, redes locales de empresa, Clouds privados o plataformas distribuidas de cómputo. Esta línea de investigación plantea un análisis profundo de todas estas variantes de arquitecturas distribuidas de bajo costo. A diferencia del análisis propuesto, las empresas proveedoras de servicios Clouds ofrecen soluciones adecuadas a costa de una infraestructura sobredimensionada en software y hardware. Diversas organizaciones pequeñas o medianas necesitan procesar sus datos en sus propios servidores. Estas organizaciones no cuentan con recursos ilimitados y la eficiencia de los mismos es el problema que trata de resolver el presente proyecto.

Palabras clave: *Distributed Computing, Cloud Computing, Cluster, SaaS, PaaS*

Contexto

El presente trabajo se encuadra dentro del área de I/D Procesamiento Distribuido y Paralelo y se enmarca dentro del proyecto de investigación: Evaluación de arquitecturas distribuidas de bajo costo basadas en software libre, que ha sido presentado oportunamente para el período 2015-2016 y está en proceso de evaluación. A pesar de ello se viene trabajando en proyectos afines que tuvieron como unidades ejecutoras al Departamento e Instituto de Informática de la FCEFYN de la UNSJ. El grupo de investigación viene trabajando desde hace más de 16 años con numerosas publicaciones y formación de recursos humanos en el área.

Introducción

La computación distribuida es un fenómeno de un crecimiento notable en los últimos años. La reducción de costos de hardware, los avances en tecnología de comunicación, el crecimiento explosivo de Internet y la dependencia cada vez más creciente de aplicaciones de red que van desde comunicación social hasta aplicaciones financieras, han contribuido a este crecimiento. El surgimiento de los sistemas embebidos, la nanotecnología y los sistemas inalámbricos, han abierto una nueva frontera para aplicaciones como

Internet de las Cosas, computación móvil, entre otras. La rápida adopción de Cloud Computing y la importancia creciente de big data han cambiado el panorama que presenta la computación distribuida.

Su aparición se debe a la necesidad de resolver problemas demasiado grandes para cualquier computadora, manteniendo la flexibilidad de trabajar en múltiples problemas más pequeños [1].

Los sistemas distribuidos proporcionan de forma transparente la compartición de recursos sean locales o remotos, facilitando el acceso y la gestión, e incrementando la eficiencia y la disponibilidad, a un costo más asequible.

Son sistemas en los cuales componentes conectados a través de una red de computadoras se comunican y coordinan sus acciones mediante el intercambio de mensajes[2]. Es una colección de computadoras independientes que dan la apariencia al usuario de ser una computadora única [3].

Existen diferentes tipos de sistemas distribuidos: cómputo distribuido, almacenamiento distribuido y sistemas ubicuos distribuidos. En el proyecto marco de este trabajo, se enfocaran las investigaciones en los sistemas de cómputo distribuidos, los cuales permiten realizar de manera más eficiente tareas de computación de alta prestaciones basadas en el paradigma de memoria distribuida. Ejemplos de Arquitecturas que soportan este tipo de sistemas son los cluster y el cloud computing.

El término cluster [4] se aplica a los conjuntos o conglomerados de computadoras construidos mediante la utilización de componentes de hardware comunes y que se comportan como si fuesen una única computadora. Hoy en día desempeñan un papel importante en la solución de problemas de las ciencias, las ingenierías y del comercio moderno.

La tecnología de clusters ha evolucionado en apoyo de actividades que van desde aplicaciones de super-cómputo y software de misiones críticas, servidores web y comercio electrónico, hasta bases de datos de alto rendimiento, entre otros usos.

El cómputo con clusters surge como resultado de la disponibilidad de procesadores económicos de alto rendimiento, redes de alta velocidad, el desarrollo de herramientas de software para cómputo distribuido de alto rendimiento, además es impulsado por la creciente necesidad de potencia computacional para variadas aplicaciones. Por su parte cloud computing es un modelo que permite el acceso en red omnipresente, conveniente y bajo demanda de un conjunto compartido de recursos informáticos configurables tales como, redes, servidores, almacenamiento, aplicaciones y servicios que pueden ser rápidamente aprovisionados y liberados con una mínima gestión esfuerzo o interacción con el proveedor o administrador de servicio.

El NIST ha presentado una de las definiciones de Cloud más aceptadas. La define como “un modelo tecnológico que permite el acceso ubicuo, adaptado y bajo demanda en red a un conjunto compartido de recursos de computación configurables (por ejemplo: redes, servidores, equipos de almacenamiento, aplicaciones y servicios), que pueden ser rápidamente aprovisionados y liberados con un esfuerzo de gestión reducido o interacción mínima con el proveedor del servicio” [5] Según dicha definición, Cloud Computing se entiende como un modelo de prestación de servicios informáticos cuya principal orientación es la escalabilidad. Desde el punto de vista de los usuarios, los servicios son elásticos, es decir, que pueden crecer o recuperar su tamaño original de manera rápida y sencilla.

Cloud está organizada en una arquitectura de capas llamada Modelo de Servicio, donde SaaS (Software as a Service) se encuentra en el tope de la pila y permite a los usuarios tener acceso a las aplicaciones alojadas en el cloud (por ej.: Google App). PaaS (Plataforma as a Service) es una capa de abstracción o middleware que se encuentra entre el SaaS y la infraestructura virtualizada, y su objetivo es proveer a los desarrolladores un framework y un conjunto de API's que puedan usar para construir sus aplicaciones (por ej.: Google App Engine). Por último, IaaS (Infraestructura as a Service) ofrece una infraestructura virtualizada de recursos (procesador, red, base de datos, disco, etc.) como servicio.

La Virtualización que se ubica como un middleware sobre el IaaS y es la encargada del aislamiento y la flexibilidad en el manejo de recursos. Gracias a la misma, es posible compartir recursos, permitiendo la asignación y reasignación de “instancias” de cada recurso. Para lograr esto es necesario contar con mecanismos de “reservación de recursos” y de “administración de recursos” para permitir el manejo consistente de todas las instancias de los recursos virtualizados con el objeto de lograr su máxima performance.

Por otro lado, se encuentra el Modelo de Despliegue, que se refieren a la posición (localización) y administración (gestión) de la infraestructura de cloud y pueden ser públicos (manejadas por terceras partes), comunitarios (operada por varias organizaciones para uso compartido) o privados (propiedad de una organización), como una cuarta opción aparece el modelo híbrido como la combinación de cualquiera de los tres modelos anteriores.

Es de suma importancia analizar cual despliegue se usará, pues esta elección impacta directamente en la performance de la solución. Hay que recordar que

cualquier elección que implique el uso intensivo de comunicaciones degrada el desempeño por los retardos incurridos en el proceso de comunicación en si [6] [7].

También es importante destacar que es posible la construcción de arquitecturas híbridas que permitan montar cluster como servicio (CaaS), permitiendo un mejor aprovechamiento de los recursos virtualizados, con el objeto de utilizar ecosistemas, como Hadoop MapReduce para procesamiento distribuido de grandes volúmenes de datos.

Tomando como base estas arquitecturas es posible convertirlas en una opción de commodities usando versiones basadas en equipamiento con recursos software y hardware limitado y conectividad tradicional. De esta manera será posible evaluar el comportamiento de estas arquitecturas en función de los diferentes algoritmos que se corran sobre ellas.

Líneas de Investigación, Desarrollo e Innovación

La mayoría de las aplicaciones de computación distribuida se centran alrededor de un conjunto de sub problemas principales. El conocimiento de estos sub problemas requiere de una base de la teoría subyacente y de numerosos aspectos algorítmicos.

Los algoritmos de álgebra lineal, al igual que los algoritmos de reparto de carga, son utilizados en computación distribuida y paralela para evaluar la performance de la distribución de datos y el cómputo. La variedad de este tipo de algoritmos permite calcular la eficiencia de los mismos para distintos niveles de granularidad y tráfico de datos.

Las organizaciones de tamaño medio o pequeño (organismos descentralizados del estado, PYMES, departamentos e institutos) necesitan mantener sus datos

privados dentro de la institución. Por lo tanto el modelo híbrido (datos en el propio centro de datos y el resto en la nube), adquiere mayor aceptación.

A pesar de la popularidad de la computación distribuida, existen escasos trabajos en áreas que asocien cloud computing con Cluster, en un sentido eficiente, para propósitos de centro de datos pequeños, cloud privados y otros sistemas distribuidos de bajo costo.

La computación distribuida presenta soporte para HPC. Esta última tiene dos líneas, que son: simulación numérica y big data, y se aplicarán algoritmos para ambos líneas como parte de las evaluaciones de la presente propuesta de proyecto de investigación.

Evaluar, analizar y construir una arquitectura de bajo costo que sea eficiente en tiempo y uso de recursos, no es tarea sencilla. La complejidad de una arquitectura distribuida está dada por la diversidad de componentes que la misma incluye como son: almacenamiento, servidores, arquitecturas de software en capas, variedad de middlewares y redes, que deben trabajar conjuntamente de la forma más óptima posible y además deben ser configurados adecuadamente para ser utilizados bajo demanda.

Los diferentes Sistemas Distribuidos presentan infraestructura que está caracterizada por diferentes atributos tales como precio, performance, robustez, consumo de energía (green computing), y por lo tanto las tareas de programación y reparto de tareas se hacen más complejos y desafiantes. Esto se complica aún más cuando se trabaja sobre supuestos de distribución de tareas multicriterio.

Algunas aplicaciones son altamente escalables y pueden correr eficientemente en ambientes de Cloud estándar, mientras que otras son dependientes de una comunicación rápida entre nodos de un Cluster dedicado y posiblemente una

arquitectura de HPC en el medio. Además la latencia afecta notablemente la escalabilidad y otros parámetros.

Algunos problemas a resolver son inherentes a los sistemas distribuidos como los requerimientos de escalabilidad, fiabilidad y disponibilidad. Al menos, parte de alguno de estos sistemas deben ser locales, puesto que existe una diferencia de microsegundos para datos locales a milisegundos para datos remotos. Por otro lado, algunas tareas se benefician de la no localidad. Por ej., tener datos duplicados en otro lugar permite la continuidad del negocio cuando un data center sale de línea.

Se deben redefinir los modelos de servicios de MapReduce en Cloud para proveer servicios de bajo costo, en contraste con el servicio tal como el que ofrece un Cloud genérico de cómputo o un Cloud de MapReduce dedicado. El mismo debe manejar adecuadamente la carga de trabajo, que tiene una significativa cantidad de trabajos interactivos, pero además presenta recursos limitados.

Los sistemas distribuidos de alto costo, suelen tener recursos casi ilimitados que incluyen sistemas de archivos de altas prestaciones como Lustre, PanFS o General Parallel File Systems, redes de muy alta velocidad como InfiniBand o Ethernet a 10Gbps, soporte para HPC con placas Nvidia Tesla o similar, administradores de recursos distribuidos como Terascale Open-source Resource and QUEUE Manager, librerías específicas para simulación científica como BT Life Science HPC o Computational Fluid Dynamic. Estas alternativas no son soportadas por sistemas distribuidos de commodity ya sea por falta de recursos de hardware o costos.

Resulta por lo tanto, necesario evaluar las diferentes plataformas de Cloud y/o Cluster adecuadas para Cloud privados o

centro de datos que resuelvan eficientemente la optimización de los recursos, para sistemas distribuidos de bajo costo con limitaciones de recursos y basados en componentes de Software libre.

Resultados y Objetivos

Resultados Obtenidos

Durante los últimos cinco años se trabajó en proyectos sobre Cloud Computing y en particular durante los últimos dos años sobre Cloud híbridos. La experiencia sobre los Cloud privados, junto con líneas de investigación anteriores [8], impulsó esta línea de investigación. El grupo ha realizado nueve publicaciones en el área durante el último año: tres trabajos en el WICC 2015, un trabajo en el CACIC 2015, dos trabajos en las Jornadas de Cloud Computing, además se realizaron tres publicaciones en revistas científicas. Se han aprobado tres tesinas de grado y un trabajo de especialización.

Objetivos

El objetivo del grupo de investigación es realizar la evaluación del desempeño y otros parámetros en arquitecturas distribuidas de commodity, analizando además como impacta la virtualización y la migración de aplicaciones y servicios entre plataformas.

Formación de Recursos Humanos

El equipo de trabajo está compuesto por los seis docentes-investigadores que figuran en este trabajo y cuatro alumnos. Se están realizando cuatro tesinas de licenciatura una sobre evaluación de algoritmos de algebra lineal sobre arquitecturas diversas, otra sobre Cloud Computing Privado, otra sobre dispositivos de juegos aplicados a salud y

otra sobre SOA aplicada a Cloud. Se espera realizar también una tesis de maestría sobre Metodologías de desarrollo aplicadas a SaaS, otra sobre bases de datos NoSQL y otra sobre algoritmos de Cómputo Intensivo para Big Data y su implementación en Clouds. Además aumentar el número de publicaciones. Por otro lado también se prevé la divulgación de varios temas investigados por medio de cursos de postgrado y actualización o publicaciones de divulgación.

Referencias

- [1] M. L. Liu – Distributed Computing Principles and Applications – Pearson Addison Wesley. (2004).
- [2] George Coulouris, Jean Dollimore, Tim Kindberg y Gordon Blair - Distributed Systems Concept and Design 5a Edition - Pearson Educación. (2012).
- [3] A. Tanenbaum, M. Van Steen – Sistemas Distribuidos Principios y Paradigmas – 2ª Edición – Pearson Prentice Hall. (2008).
- [4] R. Buyya - High Performance Cluster Computing. New Jersey: F'rentice. (1999).
- [5] Mell, P., Grance, T.. "The NIST definition of cloud computing". NIST Special Publication 800 – 145 (2011).
- [6] Murazzo, M., Rodríguez, N., Medel, D., Chávez, S., Martin, A., Valenzuela, F. "Análisis de mejora en la escalabilidad de las infraestructuras de cloud computing". III JCC & big data. UNLP (2015).
- [7] Murazzo, M., Tinetti, F., Rodríguez, N. "Despliegue de una Infraestructura Cloud Privada de Código Abierto". III JCC & big data UNLP (2015).
- [8] Rodríguez N., Murazzo M., Villafaña D., Alves M., Medel D. "Integración de Computación Heterogénea con Hadoop para Cloud Computing". WICC 2013. UADER. Paraná (2013)

Paralelización de Aplicaciones Econométricas que requieren Estimación de los Modelos de Elección Discreta

Mariano Trigila, Ricardo Di Pasquale

Ingeniería en Informática / Facultad de Ciencias Fisicomatemáticas e Ingeniería /
Pontificia Universidad Católica Argentina
Alicia Moreau de Justo 1600 (Edificio San Jose) 2° Piso, Sector Investigadores, Box 39, Buenos
Aires, Argentina
marianotrigila@gmail.com
ricardo.dipasquale@gmail.com

Resumen

En las últimas dos décadas, el uso eficiente del hardware para aplicaciones científicas fue creciendo en dificultad. Además, muchas de estas aplicaciones requieren mejorar el rendimiento del procesamiento y tratar datos masivos. Estos son sistemas complejos de implementar y en especial para aquellos que no son especialista en computación. Es necesario desarrollar e implementar abstracciones de programación de alto nivel que permitan modelos de programación simples de usar. Las aplicaciones econométricas que requieren modelos de elección discreta son aplicaciones de este tipo. En este proyecto, se aplicarán las metodologías para crear abstracciones de programación de alto nivel para extender el *framework* R, para estas aplicaciones. Los objetivos del proyecto son cubrir las necesidades de: 1) procesamiento y soporte de resolución de problemas computacionales en econometría de otros grupos de investigación; y 2) transferencia del conocimiento de las tecnologías existentes, del diseño y desarrollo de aplicaciones en el área del procesamiento paralelo para la formación de recursos humanos; y para la actualización de

graduados en informática y de otras áreas en este tipo de procesamiento que requieren las nuevas aplicaciones de gran escala. Los primeros resultados de este proyecto que comenzó, en junio de 2015, son dos cursos sobre computación paralela.

Palabras clave: econometría, computación paralela, clústeres de bajo costo.

Contexto

A partir de la finalización del proyecto de instalación del clúster de computadoras para procesamiento paralelo en el laboratorio LAPPa de la Facultad de Fisicomatemática e Ingeniería de la Pontificia Universidad Católica Argentina (UCA), el presente proyecto da comienzo a la investigación con énfasis en la integración de computación de alto rendimiento (HPC: High Performance Computing), la investigación y desarrollo en el área de algoritmos aplicados para brindar la formación y asistencia de procesamiento paralelo que requiere actualmente la comunidad de investigación y enseñanza de la UCA y de otras instituciones. Este proyecto se centra en paralelización de aplicaciones econométricas. El principal resultado de

este proyecto será la formación de recursos humanos para la investigación, la enseñanza y la asistencia a otros grupos de investigación de la Facultad de Fisicomatemáticas e Ingeniería de la UCA y de la comunidad.

Introducción

El primer desafío, y el más controvertido para el uso de la computación paralela, consiste en que la programación secuencial sigue siendo la habilidad más habitual de los programadores actuales. El segundo desafío son los sistemas operativos y compiladores. Estos fueron creciendo y evolucionaron resistentes al cambio que supone incluir el paralelismo. El tercer desafío final es como medir el rendimiento en lenguajes y frameworks de programación paralela sin que dependa estrictamente del programador o de sugerencias de otros que realizaron trabajos similares. [1].

Frente a esta situación de interrogantes, Asanovic et al. [1] de la Universidad de Berkeley plantean dos preguntas: ¿Cómo se puede desarrollar tecnología para beneficiar a los programadores actuales de la revolución paralela? y ¿Cómo se puede innovar rápidamente sin compiladores y sistemas operativos apropiados? El problema se agrava aún más si el investigador no es del área de computación.

Por lo tanto, varios laboratorios de investigación en computación paralela decidieron comenzar con la paralelización de aplicaciones para clientes [1] [2], [3], [4], [5], [6]. El enfoque es empezar con las aplicaciones, luego el software y finalmente el hardware [1]. El objetivo es encontrar una abstracción de más alto nivel para razonar acerca de los requerimientos de una aplicación paralela. La idea es definir los requerimientos de

una aplicación de tal forma que no sea excesivamente dependiente del tipo o clase de aplicación a la cual pertenece. En la Universidad de Berkeley [7], proponen que se definan métodos para capturar los requerimientos comunes de clases de aplicaciones razonablemente separadas de implementaciones individuales.

Idealmente, el objetivo es obtener buen rendimiento en el procesamiento paralelo de las aplicaciones del futuro tanto en arquitecturas con múltiples computadoras como en arquitecturas multinúcleos.

Para seleccionar las aplicaciones se sugieren cinco criterios [1]:

1. que sea desafiante en términos de impacto social y de uso en la actividad profesional;
2. que sea factible a corto plazo y que tenga potencial a largo plazo;
3. que requiera speedup significativo o moderado sobre una plataforma más eficiente y adecuada al trabajo que debe realizarse;
4. que cubra las posibles plataformas y productos que son habituales para ese tipo de problema;
5. que genere nuevas tecnologías para otras aplicaciones;
6. que involucre a expertos del área de la aplicación en el diseño, uso y evaluación de la tecnología desarrollada.

Una de las clases de aplicaciones propuesta corresponde a las aplicaciones “vergonzosamente paralelas” (*embarrassingly parallel*) [7]. Las tareas de análisis y simulación estadística en áreas tales como bioingeniería y econometría son de este tipo [8]. En particular, la Universidad de Berkeley centra el estudio sobre una de las aplicaciones representativas que es *Monte Carlo* [9]. Esta clase de aplicaciones paralelas, en particular en el

área de econometría, son requeridas por otros grupos investigación en la UCA, para su aplicación en educación y para asistir a la comunidad en general.

Estas aplicaciones tienen como característica que los cálculos dependen de resultados estadísticos de múltiples repeticiones de ensayos y las comunicaciones entre subtareas paralelas no son dominantes. Estas aplicaciones vergonzosamente paralelas son consideradas fáciles de paralelizar sobre múltiples computadoras o múltiples núcleos de una computadora, sin embargo no son problemas paralelos clásicos. Por esta razón necesitan herramientas especiales y en general, estudios de rendimiento (*speedup*) y escalamiento específicos para encontrar una paralelización eficiente que reduzca el tiempo necesario para realizar los análisis [10]. Por consiguiente, el escaso uso de computación paralela se puede explicar por las habilidades informáticas que se requieren para desarrollar este tipo de aplicaciones paralelas.

Actualmente, existen versiones paralelas de software comúnmente usado para econometría, por ejemplo, la versión *Stata/MP* de *Stata*, *Parallel Computing Toolbox* de *Matlab*, o *MP Connect Setup* de *SAS*. Cada uno de estos productos paralelos publicita que los programas estadísticos pueden ser implementados en paralelo con muy pocas modificaciones de sus métodos existentes y que reducen el tiempo de procesamiento drásticamente en comparación con la versión secuencial estándar. Las versiones paralelas de estos productos y sus licencias de uso son costosas y no suelen estar disponibles en universidades y centros de investigación.

En la literatura reciente hay múltiples artículos que demuestran el creciente interés en la computación paralela para el área de estadística y en especial usando R

[11], [12]. R es un proyecto de software libre que se distribuye bajo la licencia GNU GPL y está disponible para todos los sistemas operativos de PC. R proporciona una amplia gama de herramientas estadísticas, permitiendo que el desarrollador las extienda definiendo sus propias funciones en R e implemente bibliotecas de algoritmos más complejos de carga dinámica usando C, C++ y Fortran. R puede integrarse con distintas bases de datos y tiene una capacidad gráfica importante [13]. Sin embargo, R por sí mismo no permite ejecución en paralelo, sino que requiere de alguno de los paquetes existentes que permitan distribuir cálculos sobre múltiples computadoras o de su integración con plataformas Big Data como Apache Spark.

En el reporte técnico de Schmidberger et al. [14], se reseñan dieciséis diferentes paquetes que proveen soporte paralelo para R y se compara su estado de desarrollo, la tecnología paralela empleada, usabilidad, aceptación y rendimiento. El problema que subsiste es que requieren que el usuario/desarrollador realice el armado, configuración y administración del clúster para resolver un problema en particular. Además, la generalización es compleja ya que no todos los paquetes corren en cualquier configuración de hardware y sistema operativo, requieren programación en *MPI* o *PVM* para correr sobre un clúster y otros son específicos para *grid*.

Líneas de Investigación, Desarrollo e Innovación

El objetivo de este proyecto es la formación de recursos humanos en el área de procesamiento paralelo, para asistencia a otros grupos de investigación y su aplicación a la enseñanza de grado y

posgrado Se toma como meta la investigación en paralelización de bibliotecas del lenguaje R para aplicaciones econométricas de modelos de elección discreta con el enfoque que invita adoptar la Universidad de Berkeley [7], [1] y otros centros de computación paralela [6], [5].

Además, el tema es apropiado para la formación de recursos humanos de un proyecto conformado por investigadores que se están formando en investigación y becarios porque cumple con las líneas de investigación general, brinda ejemplos y permite estudiar exhaustivamente características, limitaciones y soluciones específicas que requieren generalización.

Las líneas de investigación son:

1. El estudio de las características de las aplicaciones econométricas y sus desafíos en la generalización. Se estudiarán los modelos de regresión logística multinomial en el contexto de modelos de elección discreta usando el framework R.
2. El estudio de las características de la computación paralela, su impacto sobre el *speedup*, sus dificultades para su generalización y los límites que imponen las plataformas de hardware y software sobre las soluciones a problemas econométricos del tipo “vergonzosamente paralelo”. En este proyecto, el estudio se centra sobre clústeres del tipo *Beowulf* y NOW (*Network of Workstations*).
3. El estudio de las metodologías de diseño para crear soluciones y abstracciones exitosas en sistemas distribuidos paralelos.
4. El desarrollo de nuevas bibliotecas y extensiones de métodos para el proyecto R. El aporte se centra en bibliotecas para aplicaciones que requieren modelos de regresión logística multinomial que usan los principios de maximización, tales

como estimación de máxima verosimilitud.

Resultados y Objetivos.

El proyecto se inició en junio de 2015. Se publicaron dos artículos en congresos internacionales como resultado de los estudios realizados sobre computación paralela para clústeres de bajo costo y clústeres NOW. Estos artículos corresponden a los resultados de instalación del clúster *Beowulf* LAPPa de la UCA y su adecuación para este proyecto [15]. El segundo artículo corresponde a un curso de postgrado para graduados en carreras de computación. Es un curso de diseño y programación de aplicaciones en diferentes ambientes paralelos (MPI, Apache Hadoop y Spark) [16].

Además, se han dictado cursos de extensión a la comunidad sobre MPI y OPENMP, cumpliendo con uno de los objetivos principales del proyecto que es la transferencia al área educativa para la formación de profesionales en el área.

En la línea de investigación correspondiente a las aplicaciones econométricas, se empezará a trabajar con asistencia de un grupo de investigación que usa econometría para su proyecto y que será el usuario inicial de las extensiones al proyecto R. Simultáneamente, se continúa investigando en el área de computación paralela.

Una vez caracterizados el ambiente de desarrollo y las aplicaciones econométricas se comenzará con el estudio de la metodología de diseño y de abstracción, para la implementación de las bibliotecas correspondientes.

Formación de Recursos Humanos

El equipo actual consta de dos investigadores. Uno de ellos está realizando su doctorado en otra institución.

En la Facultad de Ciencias Físicomatemáticas e Ingeniería, durante 2015, se desarrollaron Trabajos Finales (Tesinas) de alumnos de grado de Ingeniería en Informática en temas relacionados con las temáticas del proyecto. Como resultado se ha completado una tesina y hay 2 tesinas actualmente en curso.

Referencias

- [1] Asanovic, K., Bodik, R., Demmel, J., Keaveny, T., Keutzer, K., Kubiawicz, J., Morgan, N., Patterson, D., Sen, K., Wawrzynek, J., Wessel, D. and Yelick, K. 2009. "A View of the Parallel Computing Landscape". *Commun. ACM*, 52, 10 (Oct. 2009), 56–67.
- [2] Oxford University, Department of Statistics, <http://www.stats.ox.ac.uk/research> (1/15/2016).
- [3] Edinburg University School of Mathematics, <http://www.maths.ed.ac.uk/research> (1/15/2016)
- [4] UCLA University, Idre (Institute for Digital Research and Education) <http://www.hoffman2.idre.ucla.edu/> (1/15/2016).
- [5] Pacific Northwest National Laboratory, U.S. Department of Energy, High-Performance Computing: <http://hpc.pnnl.gov/>. (1/15/2016).
- [6]: A Strategy for Research and Innovation through High Performance Computing, Editors Mark Sawyer, Business Development and Project Manager, EPCC Mark Parsons, Executive Director, EPCC; Associate Dean for e-Research, University of Edinburg, PlanetHPC is supported under Objective "Computing Systems" of Challenge 3 "Components and Systems" of the ICT Programme of the European Commission, <http://cordis.europa.eu/fp7/ict/computing/documents/planethpc-strategy.pdf> (1/15/2016)
- [7] Asanovic, K., Bodik, R., Catanzaro, B. C., Gebis, J. J., Husbands, P., Keutzer, K., Patterson, D. A., Plishker, W.L., Shalf, J., Williams, S. W. and Yelick, K.A. 2006. *The Landscape of Parallel Computing Research: A View from Berkeley*. Technical Report #UCB/EECS-2006-183. EECS Department, University of California, Berkeley.
- [8] Grama, A., Karypis, G., Kumar, V., and Gupta A. 2003 *Introduction to Parallel Computing*. Pearson Education, second edition.
- [9] Aspuru-Guzik A, Salomon-Ferrer, R., Austin, B., Perusquia-Flores, R., Griffin, M.A., Oliva, R.A., Skinner, D., Domin, D. and Lester, Jr., W.A. 2005. "Zori 1.0: A Parallel Quantum Monte Carlo Electronic Package," *Journal of Computational Chemistry*, 26, 8, (Jun. 2005), 856–862.
- [10] Bischl, B., Lang, M., Mersmann, O., Rahnenführer, J., Weihs, C., 2012, *Computing on high performance clusters with R: Packages BatchJobs and BatchExperiments*, Technical Report, TU Technische Universität Dortmund.
- [11] Eugster, M. J. A., Knaus, J., Porzelius, C., Schmidberger, M. & Vicedo, E. 2011, Hands-on tutorial for parallel computing with R, *Computational Statistics* 26, 219–239.
- [12] Hayfield, T. & Racine, J. S. 2012, *Parallel nonparametric kernel smoothing methods for mixed data types*, <http://cran.r-project.org/web/packages/npRmpi/npRmpi.pdf> .
- [13] Project R. <https://www.r-project.org/> (2/29/2016)
- [14] Schmidberger M, Morgan M, Eddelbuettel D, Yu H, Tierney L, Mansmann U. 2009. "State of the Art in Parallel Computing with R." *Journal of Statistical Software*, 31, 1, 1–27. <http://www.jstatsoft.org/v31/i01/> (1/15/2016).
- [15] Trigila, M., 2015, "Cluster Lappa: Implementation of a Low Cost Cluster For Parallel Processing". *WCSEIT'2015 II World Congress on Systems Engineering and Information Technologies*. Vigo, España, 13. https://issuu.com/council.copec/docs/ba_wcseit2015_full (2/29/2016).
- [16] Trigila, M., Di Pasquale, R., 2016, "Teaching Parallel Computing with Low-Cost Cluster". *INTERTECH 2016: XIV International Conference on Engineering and Technology Education*, Salvador, Bahía, Brasil,

Cluster de microprocesadores RISC para problemas de agricultura de precisión

Horacio Martinez del Pezzo, Fernando Emmanuel Frati, Emmanuel Robador,
Cristian Rios

Departamento de Ciencias Básicas y Tecnológicas, Universidad Nacional de Chilecito
9 de Julio 22, Chilecito, La Rioja, Argentina
{hmartinezdelpezzo, fefrati, erobador}@undec.edu.ar,
riosbourne555@gmail.com

Resumen

En la agricultura de precisión se utilizan técnicas de cálculo estadístico y modelos de predicción climática que requieren de gran capacidad de cálculo.

Sin embargo, el costo asociado a la adquisición de un cluster de altas prestaciones resulta en muchos casos demasiado elevado. Una alternativa consiste en utilizar para tareas de cómputo intensivo una GPGPU, pero se limita a problemas de paralelismo de datos.

Este trabajo presenta una línea de I+D centrada en el desarrollo de un cluster de bajo costo destinado específicamente a resolver problemas de paralelismo de tareas.

Los temas abordados son transversales a varias áreas, como programación, electrónica digital, arquitectura de computadoras y sistemas paralelos.

Palabras clave: *cluster, ARM, RISC, sistemas paralelos*

Contexto

La línea de investigación presentada es parte del proyecto “Implementación de un

cluster de procesadores RISC” aprobado en el año 2013 por la Secretaría de Ciencia y Tecnología, convocatoria para estímulo y desarrollo de la investigación científica y tecnológica (FICyT - UNdeC). Además, cabe destacar que uno de los integrantes fue beneficiado con una beca EVC (Estímulo a las Vocaciones Científicas - Área Tecnológica, convocatoria 2015) otorgada por el CIN (Consejo Interuniversitario Nacional) para realizar tareas de formación en el marco del mencionado proyecto.

Introducción

Una característica de las regiones semi-áridas como la provincia de La Rioja es que la única agricultura sustentable es la de precisión. Para llevar adelante este tipo de agricultura es necesario aplicar técnicas de cálculo estadístico y modelos de predicción climática que requieren de gran capacidad de cálculo, dando origen a problemas de cómputo con Requerimientos de Altas Prestaciones (HPC). Normalmente, los programas para resolver estos problemas son ejecutados en clusters de altas prestaciones.

Debido a que el costo asociado a la adquisición de un cluster dedicado y su man-

tenimiento resulta muy elevado, una alternativa aceptable que ha dado muy buenos resultados consiste en configurar el cluster con un conjunto de computadoras tradicionales de bajo costo. Sin embargo, estos clusters basados en procesadores tradicionales (arquitectura CISC 80x86) enfrentan nuevos desafíos relacionados a la reducción del consumo energético y excesiva generación de calor que producen [1, 2]. Una tendencia actual para aumentar la potencia de cómputo de los clusters sin incrementar proporcionalmente el consumo consiste en equiparlos con GPGPUs y delegar cómputo en estas nuevas arquitecturas. Las GPGPUs son placas gráficas con una gran cantidad de procesadores simples pero muy efectivos en la resolución de problemas de paralelismo de datos [3, 4, 5]. Lamentablemente, la utilidad de esta tecnología es limitada en problemas donde predomina el paralelismo de tareas.

Por otro lado, el foco de la industria tecnológica se encuentra en la reducción de las dimensiones de los dispositivos, la disminución de su consumo y la aceleración de las respuestas generadas a los usuarios. Esto es posible gracias a la evolución de los microprocesadores basados en una arquitectura estándar RISC de 32 bits, y en particular la arquitectura ARM (Advanced Risc Machines) [6]. ARM se ha constituido como la arquitectura más ampliamente utilizada en unidades producidas, con presencia en más del 90 % de los dispositivos móviles desde 2009 [7].

Esta arquitectura de procesadores ha alcanzado una gran potencia de cómputo (algunos modelos superan los 1500 MHz), e incluso comienzan a aparecer versiones con varias unidades de cómputo[8]. Actualmente muchas empresas comercializan dispositivos equipados con estos procesadores que incluyen una unidad de almacenamiento, cantidades de memoria RAM y puertos

de entrada / salida similares a una computadora tradicional, pero con un costo y relación de consumo muy inferior a esta [9].

En este sentido, interesa analizar las prestaciones de un computador de propósito dedicado para la resolución de problemas de paralelismo de tareas implementado a través de un cluster de procesadores de tecnología RISC de bajo costo[10]. Este computador permitirá experimentar con problemas de cálculo estadístico aplicados a agricultura de precisión.

Esta línea de trabajo se encuentra relacionada con el proyecto Mont-Blanc del Centro de Supercomputación de Barcelona [11, 12, 13, 14], cuyo principal objetivo es encontrar solución a la limitación del consumo de energía en el futuro de los sistemas Exascale[15] con la incorporación de procesadores de bajo consumo para dispositivos móviles.

Líneas de Investigación, Desarrollo e Innovación

- Arquitecturas RISC.
- Programación de algoritmos paralelos.
- Consumo energético.
- Modelos de programación y predicción de rendimiento en estas nuevas arquitecturas.

Resultados y Objetivos

Objetivos

- Formar recursos humanos en esta tecnología emergente de carácter estratégico
- Obtener un cluster de bajo costo que sirva de entorno de experimentación

- Analizar los aspectos de estas arquitecturas que podrían ser aprovechados para resolver problemas de agricultura de precisión

Resultados parciales

Durante 2014 se adquirieron cinco equipos *cubieboard 1* y un equipo *cubieboard 2* [9]. Cada equipo *cubieboard 1* posee un procesador AllWinner A10 (ARM Cortex A8), mientras que el *cubieboard 2* posee un procesador *dualcore* AllWinner A20 (ARM Cortex A7). Todos operan a una frecuencia de 1GHz, poseen 1GB DDR3 de memoria RAM y una conexión ethernet 10/100. Se eligieron estos procesadores como entorno experimental debido a que son ampliamente utilizados en el mercado de dispositivos móviles.

Los equipos *cubieboard 1* se conectaron en red con un switch dedicado y se instaló en cada uno una distribución de Linux compatible directamente en su memoria NAND. Con respecto a la *cubieboard 2*, el desempeño general de linux instalado en su SD resultó ser demasiado lento, por lo que se está trabajando en instalar una versión compatible directamente en su NAND.

Los siguientes trabajos de investigación de alumnos han sido derivados del proyecto:

- Impacto de la red de interconexión en clusters de alto rendimiento [16].
- Análisis de desempeño de arquitecturas ARM utilizadas para problemas con requerimientos de HPC [17].
- Diseño de algoritmos paralelos para clusters de procesadores RISC de bajo costo [18].

Formación de Recursos Humanos

El equipo de trabajo está formado por tres docentes y un alumno de la carrera Ingeniería en Sistemas de la UNdeC (acreditadas por CONEAU).

Los docentes forman parte de los equipos de las asignaturas *Programación I* (2do año), *Arquitecturas de computadoras I* (2do año) y *Arquitecturas paralelas* (3er año). Dos de ellos se encuentran categorizados en el programa de incentivos, uno de los cuales es doctor en Ciencias Informáticas. El alumno posee una beca EVC desde agosto de 2015 para realizar trabajos de investigación en temas derivados del proyecto.

Las asignaturas mencionadas contemplan la aprobación mediante la participación en proyectos de investigación, por lo que pueden surgir nuevos trabajos en esta línea.

Referencias

- [1] C.-h. Hsu and W.-c. Feng, "A power-aware run-time system for high-performance computing," in *Proceedings of the 2005 ACM/IEEE Conference on Supercomputing*, ser. SC '05. Washington, DC, USA: IEEE Computer Society, 2005, p. 1-. [Online]. Available: <http://dx.doi.org/10.1109/SC.2005.3>
- [2] Q. Tang, S. K. S. Gupta, and G. Varsamopoulos, "Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 11, pp. 1458–1472, Nov. 2008.

- [3] M. F. Piccoli, *Computación de alto desempeño en GPU*. La Plata: EDULP, 2011, XV Escuela Internacional de Informática, realizada durante el XVII Congreso Argentino de Ciencia de la Computación (CACIC 2011). [Online]. Available: <http://hdl.handle.net/10915/18404>
- [4] E. Montes de Oca, L. C. De Giusti, A. E. De Giusti, and M. Naiouf, "Comparación del uso de GPU y cluster de multicore en problemas con alta demanda computacional," in *XVIII Congreso Argentino de Ciencias de la Computación (CACIC)*, Oct. 2012. [Online]. Available: <http://hdl.handle.net/10915/23623>
- [5] A. G. H. Wolfmann, "Optimización de cómputo paralelo científico en un entorno híbrido multicore - MultiGPU," in *XIV Workshop de Investigadores en Ciencias de la Computación (WICC)*, 2012, eje: Procesamiento distribuido y paralelo. [Online]. Available: <http://hdl.handle.net/10915/19393>
- [6] ARM, *ARM Architecture Reference manual*, ARM, 2013. [Online]. Available: <http://infocenter.arm.com>
- [7] J. Fitzpatrick, "An interview with Steve Furber," *Commun. ACM*, vol. 54, no. 5, p. 34–39, May 2011. [Online]. Available: <http://doi.acm.org/10.1145/1941487.1941501>
- [8] G. Yeap, "Smart mobile SoCs driving the semiconductor industry: Technology trend, challenges and opportunities," in *Electron Devices Meeting (IEDM), 2013 IEEE International*, Dec. 2013, pp. 1.3.1–1.3.8.
- [9] Cubieboard, "Main boards," Commercial web site., December 2013. [Online]. Available: <http://docs.cubieboard.org/products/start>
- [10] H. Martinez del Pezzo, E. R. D. Robador, and F. E. Frati, "Cluster de microprocesadores RISC para problemas de agricultura de precisión," in *Proceedings del XVI Workshop de Investigadores en Ciencias de la Computación*, vol. XVI. Ushuaia, Tierra del Fuego: Red UNCI, May 2014, pp. 703–705, eje: Procesamiento Distribuido y Paralelo. [Online]. Available: <http://hdl.handle.net/10915/42764>
- [11] N. Rajovic, P. Carpenter, I. Gelado, N. Puzovic, A. Ramirez, and M. Valero, "Supercomputing with Commodity CPUs: Are Mobile SoCs Ready for HPC?" 2013.
- [12] A. Auweter, "High Performance Computing on ARM Hardware. The Mont-Blanc Project," 2014.
- [13] P. Carpenter, "Building supercomputers from commodity embedded chips," 2014.
- [14] F. Mantovani, "High Performance Computing based on mobile embedded processors," 2015.
- [15] P. Thibodeau, "Scientists, IT Community Await Exascale Computers," Dec. 2009. [Online]. Available: <http://www.computerworld.com/article/2550451/computer-hardware/scientists--it-community-await-exascale-computers.html>
- [16] C. E. Rios, F. E. Frati, and P. Olmedo, "Impacto de la red de interconexión en clusters de alto rendimiento (HPC)," in *III Jornadas Científicas de Estudiantes Investigadores*

(*III-JCEI*), Universidad Nacional de Chilecito, Argentina, Oct. 2014.

- [17] M. G. Pozo, F. E. Frati, and H. Martínez del Pezzo, “Análisis de desempeño de arquitecturas ARM utilizadas para problemas con requerimientos de HPC,” in *III Jornadas Científicas de Estudiantes Investigadores (III-JCEI)*, Universidad Nacional de Chilecito, Argentina, Oct. 2014.
- [18] C. E. Rios and F. E. Frati, “Diseño de algoritmos paralelos para clusters de procesadores risc de bajo costo,” in *IV Jornadas Científicas de Estudiantes Investigadores (IV-JCEI)*, Universidad Nacional de Chilecito, Argentina, Oct. 2015.

Identificación de Algoritmos de Cómputo Intensivo para Big Data y su Implementación en Clouds

Maria A. Murazzo*, Nelson R. Rodriguez*, Miguel Guevara[&], Fernando G. Tinetti[#]

*Departamento e Instituto de Informática – FCEFy N, UNSJ.

[&]Alumno avanzado de la carrera Lic. en Sistemas de Información

[#]III-LIDI, Fac. de Informática, UNLP – Investigador CIC Prov. Bs. As.

marite@unsj-cuim.edu.ar, nelson@iinfo.unsj.edu.ar,
migueljoseguevaratencio@gmail.com, fernando@info.unlp.edu.ar

Resumen

Almacenar, transferir y procesar grandes volúmenes de datos en el área que se ha denominado *Big Data* son un factor determinante y un reto para el Cómputo de Alto Rendimiento (sigla en inglés: HPC de *High Performance Computing*). Los algoritmos usados para procesar esos datos deben sacar provecho de las ventajas ofrecidas por el cómputo en la Nube (*Cloud*), mediante el uso de algoritmos que permitan agilizar/acelerar el cómputo de o con esos datos.

La conjunción de Big Data y HPC se suele enfocar en la paralelización del procesamiento mediante la distribución de los datos y la delegación del cómputo en los nodos de procesamiento de la plataforma. Estas arquitecturas de cómputo, que para el caso de la memoria distribuida eran tradicionalmente los clusters, se pueden migrar al Cloud. La migración permite montar clusters virtuales (*Cluster as a Service*) logrando un entorno auto-escalable dependiente de la carga de trabajo. Se propone la identificación y evaluación de un conjunto representativo de algoritmos usados en Big Data con énfasis en su implementación en clouds.

Palabras clave: Cloud Computing, Big

data, HPC, Cluster Computing

Contexto

Este trabajo se encuadra dentro del área de I/D “Procesamiento Distribuido y Paralelo” y en particular dentro del proyecto de investigación “Evaluación de arquitecturas distribuidas de *commodity* basadas en software libre”, el cual ha sido presentado en la última convocatoria de CICITCA, con una duración de dos años y que tiene como unidades ejecutoras al Departamento e Instituto de Informática de la FCEFyN de la UNSJ. El grupo de investigación viene trabajando en proyectos del área desde hace más de 15 años con varias publicaciones y formación de recursos humanos.

Introducción

Debido a la aparición de nuevas tecnologías, dispositivos, medios de comunicación y aplicaciones, la cantidad de datos que se produce en la actualidad aumenta exponencialmente. Se considera que el 90% de los datos existentes se han generado en los últimos dos años, esto está dando lugar a la Era del Exa y Zetta-Byte [1]. Este aumento en la cantidad de datos demanda nuevas estrategias que permitan su almacenamiento, y análisis de manera eficiente; esto conlleva un cambio

de paradigma en las arquitecturas de cómputo, los algoritmos y también los mecanismos de procesamiento.

Frente a esta problemática se ha popularizado el término Big Data [2], el cual se usa para describir grandes conjuntos de datos, los cuales exhiben las propiedades: variabilidad, variedad, valor, volumen, velocidad, y complejidad; todas ellas denotan datos multidimensionales. Estas propiedades hacen que los sistemas de cómputo convencionales sean muchas veces inapropiadas para lograr un procesamiento adecuado [3] [4].

Los sistemas de cómputo tradicionales han evolucionado a sistemas de alto rendimiento (HPC) para llevar a cabo cómputo intensivo y mejorar la velocidad de procesamiento. Estos entornos son los ideales para Big Data debido a la creciente demanda y también creciente complejidad del cómputo necesario [5].

En este aspecto el desafío se centra en cómo se aprovecha al máximo el potencial de la arquitectura física con o por el uso de algoritmos de cómputo intensivo. Para mejorar los tiempos de procesamiento en los algoritmos, se puede optar por la implementación en plataformas con mayor potencia de cálculo como las supercomputadoras, pero los costos son elevados, lo que dificulta su acceso a una gran cantidad de comunidades científicas.

Para resolver los problemas de costo, la computación distribuida es un modelo destinado a resolver problemas de cómputo masivo utilizando un gran número de computadoras organizadas sobre una de comunicaciones. De esta manera es posible compartir recursos (quizás heterogéneos), que pueden ser basados incluso en distintas plataformas, arquitecturas y lenguajes, situados en distintos lugares y pertenecientes a diferentes dominios de administración

sobre una red que utiliza estándares abiertos [6].

En función de la problemática para la cual se decide montar una arquitectura distribuida, existen diferentes tipos de sistemas distribuidos: a) de cómputo distribuido, b) de almacenamiento y c) sistemas ubicuos distribuidos. Para el caso de esta línea de investigación se enfocarán los sistemas de cómputo distribuidos, los cuales permiten realizar de manera más eficiente tareas de computación de alto rendimiento basadas en el modelo de memoria distribuida [7]. Ejemplos de arquitecturas distribuidas para cómputo intensivo son los clusters y los clouds [8].

Según [9] cloud es un modelo de prestación de servicios informáticos cuya principal orientación es la escalabilidad. Esto es, que desde el punto de vista de los usuarios, los servicios son elásticos, o sea, pueden crecer o recuperar su tamaño original de manera rápida y sencilla. Esta orientación permite que los usuarios que acceden a los servicios, perciban que todo funciona de manera simple y rápida, dando como resultado una experiencia más gratificante.

Gracias a estas características, cloud se ha convertido en un enorme repositorio de recursos computacionales, lo cual es una buena posibilidad para construir una plataforma para las aplicaciones que necesitan una gran cantidad de recursos. Esta capacidad del cloud se debe principalmente a la habilidad de escalado elástico de recursos en función de las necesidades de las aplicaciones y el presupuesto del usuario. Esta es una tecnología centrada en ofrecer cualquier recurso (bases de datos, red, procesador, etc.) y ofrecerlo como un servicio (AaaS, Anything as a Service) bajo demanda, inclusive el cómputo [10].

Sin embargo la disponibilidad de un gran número de recursos que ofrece el cloud, también está disponibles en plataformas paralelas, las cuales ofrecen performance por sobre escalabilidad y despliegue inmediato de recursos, sin la necesidad de la virtualización

que produce una degradación del desempeño. A pesar de este inconveniente cada vez más las aplicaciones de HPC están migrando al cloud debido los aspectos económicos, que permiten contar con una plataforma configurable a las necesidades de las aplicaciones, mediante la implementación de *High Performance Computing as a Service (HPCaaS)* [11].

El paradigma cloud es atractivo para las aplicaciones HPC debido a que ofrece disponibilidad inmediata de recursos, un stack de software que puede seleccionar el dueño de la aplicación, elasticidad de los recursos, abstracción del hardware mediante la virtualización, lo cual provee portabilidad, entre las más importantes. Estos beneficios hacen que usar cloud como plataforma para los algoritmos de cómputo intensivo sea una alternativa atractiva desde el punto de vista de la performance y la eficiencia.

En los ambientes distribuidos, tales como cluster y cloud, un modelo de programación adecuado es MapReduce. MapReduce es parte de Hadoop [14], y se utiliza para desarrollar aplicaciones escalables y tolerantes a fallas en el cloud. Permite a las aplicaciones trabajar con miles de nodos y petabytes de datos

La principal motivación del modelo de programación MapReduce es la delegación del cómputo intensivo en cluster físicos o virtuales (Cluster as a Service) que, mediante un sistema de ficheros distribuido, reparte la carga de trabajo, optimizando tiempo y recursos. Asimismo, facilita un patrón de desarrollo paralelo para simplificar la implementación de algoritmos de cómputo intensivo en entornos distribuidos. Este modelo puede dividir un espacio grande de problema en espacios pequeños y paralelizar la ejecución de tareas más pequeñas en estos sub espacios [12] [13].

Líneas de Investigación, Desarrollo e Innovación

En función de lo explicado anteriormente, se consideran muy necesarias las siguientes líneas de investigación:

1. Identificar algoritmos necesarios, útiles, o al menos referenciados en big data. Caracterizar estos algoritmos al menos por área de aplicación o por tipo de procesamiento.
2. Implementar al menos algunos algoritmos en clouds. Esta implementación inicialmente puede considerar un cloud como la combinación de un cluster (en hardware y procesamiento) al que se tiene acceso vía una interfaz REST (*Representational State Transfer*).
3. Agrupar/clasificar en bibliotecas los posibles algoritmos, tanto implementados como identificados. Evaluar su utilidad o área de aplicación.
4. Analizar factores que determinarán la relación costo/rendimiento y limitaciones, tanto de los algoritmos como de la o las bibliotecas propuestas.

Resultados y Objetivos

Resultados Obtenidos

Se han realizado varias publicaciones en esta línea de investigación, entre las que se destacan [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25].

Como base para para comenzar con los trabajos de investigación se ha instalado una infraestructura cloud privada basada en OpenStack sobre la cual se está configurando Hadoop. Además, se cuenta con un cluster de seis nodos con Hadoop instalado para comenzar a realizar las investigaciones.

Como mínimo, se espera contar con una caracterización específicamente de rendimiento de los algoritmos en clouds, mostrando su rendimiento post-migración a este tipo de plataforma. Por otro lado, se espera caracterizar lo más objetivamente posible el tráfico hacia y desde el cloud para utilizar los algoritmos disponibles.

Objetivos

El principal objetivo de esta línea de investigación es tratar de identificar, proponer, implementar y determinar el rendimiento (al menos de manera experimental) de un conjunto de algoritmos de cómputo intensivo que se considere de importancia para aplicaciones de Big Data y que se orienten a ser usados en arquitecturas distribuidas.

Para realizar esta investigación, se tomarán como base las arquitecturas cluster y cloud, y sus posibles combinaciones, tales como Cluster as a Service.

Formación de Recursos Humanos

El equipo de trabajo está compuesto por los dos (2) docentes-investigadores de la F.C.E.F.y N. de la U.N.S.J. y un (1) docente-investigador de la Facultad de Informática de la U.N.L.P. y tres alumnos avanzados, dos de la Licenciatura en Ciencias de la Computación, y otro de la Licenciatura en Sistemas de Información.

Durante 2015 se rindió una tesina de grado y se están realizando dos (2) más que se rendirán en 2016.

Durante el año 2015 se realizó un (1) Trabajo Integrador de la Especialización en Redes y Seguridad (UNLP). Para el año 2016 se espera realizar una (1) tesis de la Maestría en Redes (UNLP) sobre esta temática. Además, un miembro del grupo se encuentra realizando el Doctorado en Ciencias de la Informática (UNSJ) cuya tesis aborda la línea de investigación aquí presentada.

Por otro lado también se prevé la divulgación de varios temas investigados por medio de cursos de postgrado y actualización o publicaciones de divulgación.

Referencias

- [1] Acín, Bird, Boccali, Cancio, Collier, Corney, Fuhrmann, *Architectures and methodologies for future deployment of multi-site Zettabyte-Exascale data handling platforms*. In Journal of Physics: Conference Series (Vol. 664, No. 4, p. 042009). IOP Publishing. 2015.
- [2] Nyikes, Rajnai, Z. *Big data, as part of the critical infrastructure*. In Intelligent Systems and Informatics (SISY), 2015 IEEE 13th International Symposium on (pp. 217-222). IEEE. 2015
- [3] Katal, A., Wazid, M., & Goudar, R. H. (2013, August). *Big data: issues, challenges, tools and good practices*. In Contemporary Computing (IC3), 2013 Sixth International Conference on (pp. 404-409). IEEE.
- [4] Chen, Mao, Liu, Y. (2014). *Big data: a survey*. Mobile Networks and Applications, 19(2), 171-209.
- [5] Kashyap, Fewings, Davies, Morris, Green, Guest. *Big Data at HPC Wales*. arXiv preprint arXiv:1506.08907. 2015.
- [6] Kahanwal, Singh. *The distributed computing paradigms: P2P, grid, cluster, cloud, and jungle*. arXiv preprint arXiv:1311.3070. 2013.
- [7] Tanenbaum, Van Steen. *Sistemas Distribuidos, Principios y Prácticas*. Pearson Prentice Hall. 2008.
- [8] Antonopoulos, Gillam. *Cloud Computing; Principles, Systems and Applications*. Editorial Springer Science & Business Media. 2010.
- [9] Mell, Grance. *The NIST definition of cloud computing*. NIST Special Publication 800 – 145. 2011.
- [10] Chee, Franklin. *Cloud computing: technologies and strategies of the ubiquitous data center*. CRC Press. 2010.
- [11] Petcu. *On Autonomic HPC Clouds*. Proceedings of the Second International Workshop on Sustainable Ultrascale

Computing Systems (NESUS 2015), Krakow, Poland. 2015.

[12] Malik, Sangwan. (2015). *Mapreduce Algorithms Optimizes the Potential of Big Data*. International Journal of Computer Science and Mobile Computing, Vol.4 Issue.6, June – 2015.

[13] Carrera, Geyer. *Modeling the Performance of MapReduce Applications for the Cloud*. Latin American Journal of Computing Faculty of Systems Engineering National Polytechnic School Quito-Ecuador, 2(2). 2015.

[14] Hadoop. Welcome to Apache Hadoop. <http://hadoop.apache.org>.

[15] Murazzo, Rodriguez, Chavez, Valenzuela, Martin, Guevara. *Despliegue de una arquitectura de Cloud Computing híbrida Open Source*. CONAIIISI 2014.

[16] Rodriguez, Valenzuela, Murazzo, Martin, Chavez, Villafañe, González. *Cloud Computing con herramientas libres para evaluación de modelos de despliegue híbrido*. WICC 2014.

[17] Rodriguez, Murazzo, Medel, Fernandez, Gonzalez. *Evaluación de costos de comunicación en arquitecturas para computación heterogénea aplicadas a computación científica*. WICC 2014.

[18] Rodriguez, Valenzuela, Murazzo, Chavez, Martin, Villafañe, Gonzalez.

Análisis de los parámetros de performance y escalabilidad para Clouds híbridos. SBTIC 2014.

[19] Rodriguez, Murazzo, Chavez, Guevara. *Arquitectura de Cloud Computing híbrida basada en tecnología Open Source*. CACIC 2014.

[20] Martín, Chavez, Murazzo, Rodriguez, Valenzuela. *MongoDB en ambiente Cloud Híbrido con OpenStack*. WICC 2015.

[21] Rodriguez, Valenzuela, Murazzo, Chavez, Martin, Villafañe. *Estudio y Análisis de estrategias de seguridad para Cloud Computing híbridos*. WICC 2015.

[22] Murazzo, Tinetti, Rodriguez, Guevara. *Infraestructura de Cloud Computing*. WICC 2015.

[23] Murazzo, Tinetti, Rodriguez. *Despliegue de una Infraestructura Cloud Privada de Código Abierto*. III JCC & Big Data 2015.

[24] Rodriguez, Murazzo, Medel Chavez, Martin, Valenzuela. *Análisis de mejora en la escalabilidad de las infraestructuras de cloud computing*. III JCC & Big Data 2015.

[25] Murazzo, Rodriguez. *Evaluación del Impacto de Migración al Cloud*. SBTIC 2015.

Fundamentos de cómputo paralelo y distribuido para HPC. Construcción y evaluación de aplicaciones.

Marcelo Naiouf⁽¹⁾, Armando De Giusti⁽¹⁾⁽²⁾, Laura De Giusti⁽¹⁾, Franco Chichizola⁽¹⁾, Victoria Sanz⁽¹⁾⁽²⁾, Adrián Pousa⁽¹⁾, Enzo Rucci⁽¹⁾⁽²⁾, Silvana Gallo⁽¹⁾⁽²⁾, Erica Montes de Oca⁽¹⁾, Emmanuel Frati⁽¹⁾, Mariano Sánchez⁽¹⁾, María José Basgall⁽¹⁾⁽²⁾, Carolina Actis⁽¹⁾, Adriana Gaudiani⁽³⁾

¹Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – Universidad Nacional de La Plata

² CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas

³ Universidad Nacional de General Sarmiento

{mnaouf, degiusti, ldgiusti, francoch, vsanz, apousa, erucci, sgallo, emontesdeoca, fefrati, msanchez, mjbasgall. cactis}@lidi.info.unlp.edu.ar, agaudi@ungs.edu.ar

Resumen

El eje central de la línea presentada son los temas de procesamiento paralelo y distribuido para HPC (fundamentos y aplicaciones). Interesa la construcción, evaluación y optimización de soluciones con algoritmos concurrentes, paralelos y distribuidos sobre diferentes plataformas de software y arquitecturas con múltiples procesadores (multicore, clusters de multicore, cloud y aceleradores como GPU, FPGA y Xeon Phi), los lenguajes y paradigmas de programación paralela (puros e híbridos), los modelos de representación de aplicaciones paralelas, los algoritmos de (mapping y scheduling), el balance de carga, las métricas de evaluación de complejidad y rendimiento (speedup, eficiencia, escalabilidad, consumo energético), y la construcción de ambientes para la enseñanza de la programación concurrente y paralela.

Se propone aplicar los conceptos en problemas numéricos y no numéricos de cómputo intensivo y/o sobre grandes volúmenes de datos (búsquedas, simulaciones, n-body, imágenes, big data, reconocimiento de patrones, bioinformática, etc), con el fin de obtener soluciones de alto rendimiento.

En la dirección de tesis de postgrado existe colaboración con el grupo HPC4EAS (High Performance Computing for Efficient Applications and Simulation) del Dpto. de Arquitectura de Computadores y Sistemas Operativos de la Universidad Autónoma de Barcelona, y con el Departamento de Arquitectura de Computadores y Automática de la Universidad Complutense de Madrid, entre otros.

Palabras clave: Cómputo paralelo y distribuido de altas prestaciones. Algoritmos paralelos y distribuidos. Clusters. Multicore. GPU. Consumo energético. Balance de carga. Aplicaciones. Evaluación de performance.

Contexto

La línea de I/D que se presenta en este trabajo es parte del Proyecto 11/F017 “Cómputo Paralelo de Altas Prestaciones. Fundamentos y Evaluación de rendimiento en HPC.

Aplicaciones a Sistemas Inteligentes, Simulación y Tratamiento de Imágenes” del III-LIDI acreditado por el Ministerio de Educación, y de proyectos acreditados y subsidiados por la Facultad de Informática de la UNLP. Además, hay cooperación con Universidades de Argentina, Latinoamérica y Europa a través de proyectos acreditados por AECID, CyTeD, y OEI, y becas de Telefónica de Argentina. Asimismo, el Instituto forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD).

Introducción

El área de procesamiento paralelo se ha convertido en clave dentro de las Ciencias de la Computación, debido al creciente interés por el desarrollo de soluciones a problemas con muy alta demanda computacional y de almacenamiento, produciendo transformaciones profundas en las líneas de I/D [RAU10][PAC11][KIR12].

El desafío se centra en cómo aprovechar las prestaciones obtenidas a partir de la evolución de las arquitecturas físicas. En esta línea de I/D la mayor importancia está en los algoritmos paralelos y en los métodos utilizados para su construcción y análisis a fin de optimizarlos.

Uno de los cambios de mayor impacto ha sido el uso de manera masiva de procesadores con más de un núcleo (*multicore*), produciendo plataformas distribuidas híbridas (memoria compartida y distribuida) y generando la necesidad de desarrollar sistemas operativos, lenguajes y algoritmos que las usen adecuadamente. También creció la incorporación de placas aceleradoras a los sistemas multicore constituyendo plataformas paralelas de memoria compartida con paradigma de programación propio asociado. Asimismo, los entornos de computación cloud introducen un nuevo foco desde el punto de vista del HPC, brindando un soporte “a medida” para la ejecución de aplicaciones sin la necesidad de adquirir el hardware.

La creación de algoritmos paralelos en arquitecturas multiprocesador no es un proceso directo [MCC12]. El costo puede ser alto en términos del esfuerzo de programación y el manejo de la concurrencia adquiere un rol central en el desarrollo. Los pasos básicos para diseñar aplicaciones paralelas incluyen particionamiento, comunicación, aglomeración y mapeo de procesos a

procesadores, y si bien en las primeras etapas el diseñador puede abstraerse de la máquina sobre la que ejecutará el algoritmo, para obtener buen rendimiento debe tenerse en cuenta la plataforma de destino. En las máquinas multiprocesador, se deben identificar las capacidades de procesamiento, interconexión, sincronización y escalabilidad [PAR09]. La caracterización y estudio de rendimiento del sistema de comunicaciones es de interés para la predicción y optimización de performance, así como la homogeneidad o heterogeneidad de los procesadores.

Muchos problemas algorítmicos se vieron impactados por las máquinas multicore y el uso de clusters de multicore. A partir de incorporar varios chips multicore dentro de un nodo y conectar múltiples nodos vía red, se puede crear una arquitectura NUMA, de modo que los cores en un chip compartan memoria principal, y puedan acceder remotamente a la memoria dedicada de otro chip, aunque ese acceso sea más costoso, surgiendo así varios niveles de comunicación. Esto impacta sobre el desarrollo de algoritmos que aprovechen adecuadamente las arquitecturas, y motiva el estudio de performance en sistemas híbridos [SID07]. Además, es necesario estudiar la utilización de diferentes lenguajes ya que aún no se cuenta con un standard, aunque puede mencionarse el uso de MPI, OpenMP y Pthreads [MUR11].

Para algunos problemas ha crecido la utilización de placas aceleradoras, como pueden ser las unidades de procesamiento gráfico (GPU, graphic processing unit) de NVIDIA y AMD o los coprocesadores Xeon Phi de Intel [PIC11][KIR12][JEF13]. Esto se debe a la capacidad que tienen estos dispositivos de alcanzar picos de rendimiento y cocientes de eficiencia energética superiores a los de la CPU a un menor costo. Por otro lado, el uso de FPGAs (Field Programmable Gate Array) se ha vuelto atractivo para HPC debido a la evolución en su capacidad de cómputo, su bajo consumo energético y al desarrollo de nuevas herramientas de programación más familiares para el área [SET13][XIL15].

La combinación de arquitecturas con múltiples núcleos con aceleradores dio lugar a plataformas híbridas con diferentes características [RUC16b]. Más allá del tipo de acelerador utilizado, la programación de esta clase de plataformas representa un verdadero desafío. Para lograr aplicaciones de alto rendimiento, los programadores deben enfrentar diversas dificultades como pueden ser: estudiar características específicas de cada arquitectura y aplicar técnicas de programación y optimización particulares para cada una de ellas, lograr un balance de carga adecuado entre los diferentes dispositivos de procesamiento y afrontar la ausencia de un estándares y de herramientas avanzadas para este tipo de sistemas.

Por otra parte, los avances en las tecnologías de virtualización han dado origen al paradigma de Cloud Computing, que se presenta como una alternativa a los tradicionales sistemas de cluster [EC213][OPE13]. El uso de cloud para HPC presenta desafíos atractivos, brindando un entorno reconfigurable dinámicamente sin la necesidad

de adquirir hardware, y es una excelente plataforma para testear escalabilidad de algoritmos aunque queda mucho por hacer en cuanto al diseño, lenguajes y programación

Métricas de evaluación del rendimiento y balance de carga

La diversidad de opciones vuelve complejo el análisis de performance de los Sistemas Paralelos, ya que los ejes sobre los cuales pueden compararse dos sistemas son varios. Existe un gran número de métricas para evaluar el rendimiento, siendo las tradicionales: tiempo de ejecución, speedup, eficiencia.

La *escalabilidad* permite capturar características de un algoritmo paralelo y la arquitectura en que se lo implementa. Posibilita testear la performance de un programa sobre pocos procesadores y predecirla en un número mayor, así como caracterizar la cantidad de paralelismo inherente en un algoritmo.

En arquitecturas distribuidas, los problemas que caracterizan el análisis de los algoritmos paralelos aparecen potenciados por las dificultades propias de la interconexión en una red en general no dedicada. Esto se torna más complejo aún si cada nodo puede ser un multicore con varios niveles de memoria.

El uso de procesadores con múltiples núcleos conlleva cambios en la forma de desarrollar aplicaciones y software, y evaluar su rendimiento. La cantidad de threads disponibles en estos sistemas también es importante, ya que su creación y administración requiere del uso de recursos como memoria; además los threads deben ser cuidadosamente planificados (scheduling) e incorporados en la pila de ejecución. En este sentido, el desarrollo de técnicas eficientes es un tema de interés.

El objetivo principal del cómputo paralelo es reducir el tiempo de ejecución haciendo uso eficiente de los recursos. El *balance de carga* es un aspecto central y consiste en, dado un conjunto de tareas que comprenden un algoritmo y un conjunto de procesadores, encontrar el mapeo (asignación) de tareas a procesadores tal que cada una tenga una cantidad de trabajo que demande aproximadamente el mismo tiempo, y esto es más complejo si hay heterogeneidad. Dado que el problema general de mapping es *NP*-completo, pueden usarse enfoques que dan soluciones subóptimas aceptables [OLI08]. Las técnicas de planificación a nivel micro (dentro de cada procesador) y macro (en un cluster) deben ser capaces de obtener buen balance de carga. Existen técnicas estáticas y dinámicas cuyo uso depende del conocimiento que se tenga sobre las tareas que componen la aplicación [DUM08].

Evaluación de performance. Aplicaciones

Es importante referirse a un algoritmo paralelo mencionando el modelo de computación para el que fue diseñado. Uno de los objetivos en la definición del modelo es la posibilidad de predicción de performance que brinde el mismo, teniendo en cuenta conceptos tales como comunicación, sincronización y arquitectura. El desarrollo

de nuevos modelos requiere caracterizar el contexto de comunicaciones entre los procesadores y la asociación entre los algoritmos, el paradigma elegido y la arquitectura. En la evaluación de performance de distintas clases de aplicaciones sobre las arquitecturas, interesa estudiar la influencia de las estrategias de distribución de procesos y datos, y la carga (estática o dinámica) asignada a cada procesador sobre el speedup, eficiencia y escalabilidad. Un aspecto de interés que se ha sumado como métrica es el del consumo energético requerido [BAL12]. Entre las aplicaciones de interés se encuentran las numéricas y no numéricas con alta demanda de cómputo.

Líneas de Investigación, Desarrollo e Innovación

- Paralelización de algoritmos secuenciales. Diseño y optimización de algoritmos.
- Comparación de lenguajes y bibliotecas para procesamiento paralelo y distribuido.
- Estudio de complejidad de algoritmos paralelos, considerando multicore y heterogeneidad.
- Modelos y paradigmas de computación paralela. Modelo Map-reduce.
- Modelos de representación y predicción de performance de algoritmos paralelos.
- Arquitecturas multicore y many-core. Arquitecturas FPGA.
- Multiprocesadores distribuidos.
- Arquitecturas híbridas (diferentes combinaciones de multicores y GPUs) y Arquitecturas heterogéneas.
- Programación sobre modelos híbridos: pasaje de mensajes y memoria compartida en cluster de multicores, clusters de GPU, clusters multicore-GPU
- Técnicas de programación sobre arquitecturas many-core (GPU y Xeon Phi) y FPGA.
- Técnicas para soluciones de HPC en cloud.
- Lenguajes y Estructuras de Datos para nuevas arquitecturas de cómputo paralelo.
- Mapping y scheduling de aplicaciones paralelas sobre distintas arquitecturas multiprocesador. Balance de carga estático y dinámico. Técnicas.
- Análisis de los problemas de migración y asignación de procesos y datos a procesadores.
- Desarrollo de soluciones paralelas a problemas de cómputo intensivo y/o con grandes volúmenes de datos (búsquedas, simulaciones, n-body, aplicaciones científicas, big data, bioinformática) sobre diferentes modelos de arquitectura homogéneas y heterogéneas (multicores, clusters, clusters de multicore, GPU, Xeon Phi, FPGA y cloud).
- Evaluación de rendimiento, eficiencia energética y costo de programación de las diferentes soluciones implementadas teniendo en cuenta las arquitecturas y las herramientas de programación utilizadas.
- Ambientes para la enseñanza de programación concurrente

Resultados y Objetivos

- Formar RRHH en los temas del Subproyecto, incluyendo tesis de postgrado y tesinas de grado.
- Desarrollar y optimizar algoritmos paralelos sobre diferentes modelos de arquitectura. En particular, en aplicaciones numéricas y no numéricas de cómputo intensivo y tratamiento de grandes volúmenes de datos (big data).
- Utilizar arquitecturas híbridas que combinan memoria compartida y pasaje de mensajes, evaluando performance para distintos modelos de comunicación.
- Estudiar y comparar los lenguajes sobre las plataformas multiprocesador para diferentes modelos de interacción entre procesos.
- Investigar la paralelización en plataformas que combinan multicore y aceleradores, o que disponen de más de un acelerador. Comparar estrategias de distribución de trabajo teniendo en cuenta las diferencias en potencias de cómputo y comunicación, dependencia de datos y memoria requerida.
- Evaluar la performance (speedup, eficiencia, escalabilidad, consumo energético) de las soluciones propuestas. Analizar el rendimiento de soluciones paralelas a problemas con diferentes características (dependencia de datos, relación cómputo / comunicación, memoria requerida).
- Emplear experimentalmente contadores de hardware orientados a la detección de fallas de concurrencia y evaluación del rendimiento.
- Mejorar y adecuar las técnicas disponibles para el balance de carga (estático y dinámico) entre procesos a las arquitecturas consideradas.

En este marco, pueden mencionarse los siguientes resultados:

- Para la experimentación se utilizaron y analizaron diferentes arquitecturas homogéneas o heterogéneas, incluyendo multicores, cluster de multicores (con 128 núcleos), GPU y cluster de GPU, Xeon Phi y FPGA.
- Se experimentó la paralelización en arquitecturas híbridas, con el objetivo de estudiar el impacto del mapeo de datos y procesos, así como de los lenguajes y librerías utilizadas.
- Respecto de los aplicaciones estudiadas y algoritmos implementados, se trabajó fundamentalmente con los siguientes problemas:

- **Best-first search paralelo sobre multicore y cluster de multicore:** El algoritmo de búsqueda A* (variante de Best-First Search) es utilizado como base para resolver problemas combinatorios y de planificación, donde se requiere encontrar una secuencia de acciones que minimicen una función objetivo para transformar una configuración inicial en una configuración final. El alto requerimiento de memoria y cómputo causados por el crecimiento exponencial del árbol de búsqueda generado dinámicamente hacen imprescindible su paralelización, que permite beneficiarse de: (a) la gran cantidad de RAM

y potencia de cómputo de un *cluster*, (b) la potencia de cómputo de los procesadores *multicore*, (c) ambas características en caso de *cluster de multicore*. Tomando como caso de estudio el problema del N-Puzzle, se implementó el algoritmo A* secuencial para resolverlo, y dos versiones propias optimizadas del algoritmo paralelo Hash Distributed A* (HDA*[KIS12] [BUR10]) que realiza el balance de carga de los nodos generados del grafo mediante una función hash: (1) una versión utiliza MPI, ejecutando tanto sobre arquitecturas con memoria distribuida y memoria compartida, y (2) otra versión con Pthreads, que ejecuta sobre un multiprocesador con memoria compartida, que elimina ciertas ineficiencias respecto a (1). Se analizó el rendimiento de ambas versiones cuando corren sobre una máquina multicore, al aumentar la cantidad de cores y la carga de trabajo, comprobando su buena escalabilidad sobre dicha arquitectura. También se comprobó que HDA* Pthreads obtiene mejor rendimiento y reduce el consumo de memoria respecto a HDA* MPI cuando corren sobre multicore, estos resultados indican que será beneficioso aplicar programación híbrida para este algoritmo de búsqueda cuando la arquitectura subyacente será un cluster de multicore y en consecuencia se sentaron las bases para el algoritmo HDA* híbrido. Se comprobó la buena escalabilidad de HDA* MPI sobre un cluster de multicore convencional. Las optimizaciones añadidas en las versiones desarrolladas mejoraron el rendimiento respecto a los algoritmos originales. Como líneas de trabajo futuro se plantea implementar el algoritmo HDA* híbrido y analizar su rendimiento sobre un cluster de multicore [SAN15].

➤ **Criptografía de grandes volúmenes de datos.** Hoy en día, el volumen de datos que se transmiten en las redes se ha incrementado considerablemente, y en ocasiones suelen ser información sensible, por lo tanto es importante codificarlos para enviarlos por una red pública como lo es InterNet de manera segura. El encriptado y desencriptado de datos requiere un tiempo de cómputo adicional, que dependiendo de su tamaño puede ser muy alto. AES (Advanced Encryption Standard), es un algoritmo de cifrado simétrico por bloques que se ha convertido en estándar en 2002, y actualmente es el más ampliamente usado para codificar información. Este algoritmo se caracteriza por ser simple, rápido y por consumir pocos recursos. Sin embargo el tiempo de cifrar y descifrar grandes cantidades de datos es importante por lo que es oportuno aprovechar las posibilidades que brindan las arquitecturas multicore para reducir este tiempo. Para este propósito, las arquitecturas multicore actuales, como clusters de multicore y GPUs, proporcionan una forma de acelerar del cómputo de encriptar y desencriptar información logrando un excelente rendimiento [POU15].

➤ **Búsqueda de similitud en bases de datos biológicas.** El algoritmo de Smith-Waterman es el método más preciso para esta clase de búsquedas, la cual resulta ser una tarea cotidiana y recurrente en las investigaciones de la bioinformática y la biología molecular.

Desafortunadamente el algoritmo Smith-Waterman resulta costoso debido a su complejidad computacional cuadrática y la situación se agrava a causa del crecimiento exponencial de datos biológicos en los últimos años [RUC16a]. El reciente surgimiento de aceleradores en HPC (GPU, Xeon Phi, FPGA, entre otros) da la oportunidad de acelerar las búsquedas biológicas sobre hardware comúnmente disponible a un costo accesible. Por ese motivo, se desarrollaron diferentes soluciones paralelas para arquitecturas heterogéneas basadas en coprocesadores Intel Xeon Phi. Se analizó el rendimiento, la eficiencia energética y el costo de programación de cada solución y se los comparó con el de otras implementaciones basadas en CPU y GPU [RUC15a]. Por otra parte, se desarrolló un primer estudio de viabilidad del empleo del modelo de programación OpenCL sobre FPGAs para acelerar este problema [RUC15b]. A futuro, interesa profundizar el estudio de soluciones basadas en FPGAs y comparar su rendimiento, costo de programación y eficiencia energética con el de las empleadas anteriormente.

➤ **Simulación distribuida de modelos orientados al individuo.** El modelado orientado al individuo resulta de gran interés ya que permite analizar y extraer conclusiones acerca de un sistema a través de la simulación de la interacción de sus individuos. No obstante, a medida que los modelos incorporan más características del sistema se vuelven más complejos y en consecuencia se necesita mayor cantidad de cómputo y comunicación para lograr resultados significativos. La aparición de arquitecturas distribuidas y procesadores con varios núcleos ha favorecido el desarrollo de dichos modelos a gran escala utilizando técnicas de simulación distribuida. Lograr mejoras en los tiempos de ejecución y en la eficiencia de dichas aplicaciones es esencial. Con el fin de mejorar la distribución de la carga entre los procesos lógicos de la simulación se pretende analizar el desempeño del simulador en cuanto al balance de cómputo, ya que al trabajar con grandes cantidades de individuos, y debido a la cantidad de cómputo asociada a la selección de vecinos para realizar los desplazamientos, se producen desbalances en cuanto a la cantidad de trabajo de cada uno de los diferentes procesos lógicos. Dichos desbalances obstaculizan el aprovechamiento de la arquitectura disponible, por lo que se deben estudiar diferentes técnicas de balance de cómputo y desarrollar las soluciones que aporten dichas mejoras.

➤ **Problemas de tipo N-body.** Se utilizaron las plataformas de memoria compartida GPU y cluster de multicore para la resolución de problemas con alta demanda computacional del tipo N-body. Se emplearon diferentes modelos de comunicación: memoria compartida (Pthreads en CPU y CUDA en GPU), pasaje de mensajes (MPI) y soluciones híbridas (MPI-Pthreads). Además, se analizó el problema desde el punto de vista energético, introduciendo los fundamentos de consumo energético. Se han mostrado los beneficios del uso de la GPU en problemas con características similares al caso planteado.

Los tiempos de ejecución obtenidos son considerablemente inferiores comparados con las soluciones implementadas en CPU. Los experimentos realizados desde el punto de vista del consumo energético favorecen el uso de la GPU en tales problemas sobre un cluster de GPU, utilizando la combinación de MPI-CUDA. El trabajo experimental ha dado como resultado una buena aceleración obtenida utilizando cluster de GPU. Además, se comparó la aceleración de la aplicación ejecutada en un cluster de CPU y en el cluster de GPU; se observó claramente que el uso de este último logró una aceleración similar al uso del cluster de CPU pero con tiempos de ejecución significativamente menores [MON14]. Actualmente, se trabaja en la medición de consumo energético sobre cluster de GPU.

➤ **Problemas de simulación relacionados con fenómenos naturales (inundaciones).** Análisis de diferentes soluciones para la paralelización de este tipo de aplicaciones que son intensivas en cómputo; y el tiempo de ejecución y la performance alcanzable son críticas dado que los resultados que se esperan determinarán alertas y toma de decisiones. La utilización de escenarios de simulación en entornos donde interesa estudiar el comportamiento en situaciones de desastres producidos por fenómenos naturales como las inundaciones. En este ámbito se avanza en dos temas: (1) La implementación de un método de sintonización de un simulador de inundaciones en ríos de llanura, mediante la técnica de simulación paramétrica. El proceso requiere lanzar miles de escenarios de simulación hasta encontrar un conjunto ajustado de parámetros de entrada del simulador. La experimentación se lleva a cabo con un modelo master-worker sobre un cluster [GAU15]. (2) En colaboración con el Laboratorio de Hidrología de la UNLP se comenzó con la paralelización de la simulación de inundaciones producidas por lluvias (en particular en el ámbito de la ciudad de La Plata, donde una corrida “standard” es del orden de las 8 hs), a fin de reducir el tiempo de ejecución a pocos minutos y permitir establecer un sistema de alertas [GAU16].

➤ **Ambientes para la enseñanza de concurrencia.** Se desarrolló el entorno R-INFO para la enseñanza de programación concurrente y paralela a partir de cursos iniciales en carreras de Informática. Incluye un entorno visual que representa una ciudad en la que pueden definirse varios robots que interactúan. Combina aspectos de memoria compartida y distribuida mediante instrucciones para bloquear y liberar esquinas de la ciudad y el concepto de pasaje de mensajes a través de primitivas de envío y recepción. Además se incluyen los conceptos de heterogeneidad (diferentes velocidades de los robots) y consumo energético [DEG15]. Se ha integrado con el uso de robots físicos (Lego Mindstorm 3.0) que ejecutan en tiempo real las mismas instrucciones que los robots virtuales y se comunican con el entorno mediante bluetooth [DEG14].

➤ **Aplicaciones en Big Data.** Actualmente, la cantidad de datos que se crea cada dos días se estima que es equivalente a lo generado desde el principio de los tiempos hasta 2003. Los datos producidos son del orden superior a los petabytes (10¹⁵ bytes), y crece en torno al 40 % cada año. Muchas de las técnicas para el tratamiento sobre “Grandes Datos” pertenecen a la minería de datos, y en este sentido es de interés estudiar la paralelización de las mismas a fin de obtener resultados en tiempos razonables. En particular, se está trabajando sobre grandes volúmenes de datos en flujo continuo provenientes de redes de microblogging, particularmente Twitter, estudiando la paralelización de las mismas bajo el paradigma MapReduce [BAS15]. Se trata de una línea incipiente en el grupo.

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D se concluyeron 2 tesis doctorales, 3 Trabajos Finales de Especialización y 2 Tesinas de Grado de Licenciatura. Se encuentran en curso en el marco del proyecto 7 tesis doctorales, 2 de maestría, 3 trabajos de Especialización y 2 Tesinas.

Se participa en el dictado de las carreras de Doctorado en Cs. Informáticas y Magíster y Especialización en Cómputo de Altas Prestaciones de la Facultad de Informática UNLP, por lo que potencialmente pueden generarse más Tesis y Trabajos Finales.

Hay cooperación con grupos de otras Universidades del país y del exterior, y tesistas de diferentes Universidades realizan su trabajo con el equipo del proyecto.

Referencias

- [BAL12] Balladini J., Rucci E., De Giusti A., Naiouf M., Suppi R., Rexachs D., Luque E. “Power Characterisation of Shared-Memory HPC Systems”. Computer Science & Technology Series – XVIII Argentine Congress of Computer Science Selected Papers. Editores: Armando De Giusti, Guillermo Simari, Patricia Pesado. ISBN 978-987-1985-20-3. Págs. 53-65. Editorial de la Universidad de La Plata (edulp). La Plata (Argentina). 2013
- [BAS15] Basgall M. J., Aquino G., Lanzarini L., Naiouf M. “Un Enfoque Dinámico para la Detección de Relaciones entre Tópicos en Textos provenientes de Redes Sociales”. III Jornadas de Cloud Computing y Big Data (JCC&BD). Facultad de Informática. UNLP. 2015.
- [BUR10] Burns E, Lemons S, Ruml W, Zhou R. “Best First Heuristic Search for Multicore Machines”. Journal of Artificial Intelligence Research, Vol.39, No.1, pp. 689-743, 2010.
- [DEG14] De Giusti L., Leibovich F., Sanchez M., Chichizola F., Naiouf M., De Giusti A. "Herramienta interactiva para la enseñanza temprana de Concurrencia y Paralelismo: un caso de estudio", Procs XX Congreso Argentino de Ciencias de la Computación – Workshop de Innovación en Educación. Octubre 2014. Pp 133-140. ISBN: 978-987-3806-05-6
- [DEG15] De Giusti L., Leibovich F., Chichizola F., Naiouf

- M., De Giusti A. "Incorporando conceptos en la enseñanza de Concurrencia y Paralelismo utilizando el entorno CMRE". Procs XXI Congreso Argentino de Ciencias de la Computación – Workshop de Innovación en Educación. Octubre 2015. Pp 1212-1221. ISBN: 978-987-3724-37-4
- [DUM08] Dummler J., Ruaber T., Runger G., Mapping Algorithms for Multiprocessor Tasks on Multi-Core Clusters, Proc. 2008 International Conference on Parallel Processing IEEE CS 2008.
- [EC213] Amazon Elastic Compute Cloud (Amazon EC2). <http://aws.amazon.com/es/ec2/>. Febrero 2013.
- [GAU15] Adriana Gaudiani. "Simulación y optimización como metodología para mejorar la calidad de la predicción en un entorno de simulación hidrográfica". Tesis de Doctorado en Ciencias Informáticas (Facultad de Informática – UNLP). 2015.
- [GAU16] Adriana Gaudiani, Emilio Luque, Pablo García, Mariano Re, Marcelo Naiouf, Armando De Giusti. "How a Computational Method Can Help to Improve the Quality of River Flood Prediction by Simulation". Advances and New Trends in Environmental and Energy Informatics (part V). ISBN 978-3-319-23455-7. Pp337-351. 2016.
- [JEF13] Jeffers, James; Reinders, James. "Intel Xeon Phi Coprocessor High Performance Programming", Morgan Kaufmann, 2013.
- [KIR12] Kirk D., Hwu W. "Programming Massively Parallel Processors, second edition: A Hands-on Approach. Morgan-Kaufmann. 2012.
- [KIS12] Kishimoto A., Fukunaga A., Botea A. "Evaluation of a Simple, Scalable, Parallel Best-FirstSearch Strategy", Arxivpreprint: arXiv 1201.3204, 2012.
- [MCC12] McCool, Michael. "Structured Parallel Programming: Patterns for Efficient Computation", Morgan Kaufmann, 2012
- [MON14] E. Montes de Oca, L. De Giusti, F. Chichizola, A. De Giusti, M. Naiouf. "Utilización de Cluster de GPU en HPC. Un caso de estudio". Proceedings del XX Congreso Argentino de Ciencias de la Computación (CACIC 2014) – Workshop de Procesamiento Distribuido y Paralelo. Octubre 2014. Pp 1220-1227. ISBN: 978-987-3806-05-6
- [MUR11] Muresano Cáceres R. "Metodología para la aplicación eficiente de aplicaciones SPMD en clústers con procesadores multicore" Ph.D. Thesis, UAB, Barcelona, España, Julio 2011.
- [NOT09] Nottingham A. y Irwin B. "GPU packet classification using opencl: a consideration of viable classification methods". Research Conf. of the South African Inst. of Comp. Sc. and Inf. Technologists. ACM, 2009.
- [OLI08] Olivier S., Prins S., Scalable Dynamic Load Balancing Using UPC. Proc. 37th ICPP'08. CD-ROM, IEEE CS, September 2008.
- [OPE13B] OpenStack Cloud Software: Open source software for building private and public clouds. <http://www.openstack.org>. Febrero 2013.
- [PAC11] Pacheco, Peter. "An Introduction to Parallel Programming". Morgan Kaufmann, ISBN-13: 978-0123742605, 2011.
- [PAR09] Parashar M., Li Xiaolin, Chandra Sumir, "Advanced Computational Infrastructures for Parallel and Distributed Applications", Wiley-Interscience, 2009 ISBN-10: 0470072946
- [PIC11] Piccoli M.F., "Computación de Alto Desempeño utilizando GPU". XV Escuela Internacional de Informática. Editorial Edulp, 2011.
- [POU15] Adrian Pousa, Victoria Sanz, Armando De Giusti. "Comparación de rendimiento de algoritmos de cómputo intensivo y de acceso intensivo a memoria sobre arquitecturas multicore. Aplicación al algoritmo de criptografía AES". XXI Congreso Argentino de Ciencias de la Computación (Junín, 2015). ISBN 978-987-3806-05-6.
- [RAU10] Rauber T., Rüniger G. "Parallel programming for multicore and cluster systems". Springer. 2010.
- [RUC15a] Enzo Rucci, Carlos García Sanchez, Guillermo Botella Juan, Armando De Giusti, Marcelo Naiouf, Manuel Prieto-Matías. "An Energy-Aware Performance Analysis of SWIMM: Smith-Waterman Implementation on Intel's Multicore and Manycore". Concurrency and Computation: Practice and Experience. Vol. 27 Nro. 18. Págs- 5517-5537. ISSN Online: 1532-0626, 1532-0634 (Online). DOI: 10.1002/cpe.3598.
- [RUC15b] Enzo Rucci, Armando De Giusti, Marcelo Naiouf, Carlos García Sanchez, Guillermo Botella Juan, Manuel Prieto-Matías. "Smith-Waterman Protein Search with OpenCL on FPGA". Proceedings of 2014 IEEE Symposium on Parallel and Distributed Processing with Applications (ISPA). 20 al 22 de Agosto de 2015. Helsinki, Finlandia. ISBN: 978-1-4673-7952-6. Págs. 208-213. DOI: 10.1109/Trustcom.2015.634.
- [RUC16a] Enzo Rucci, Armando De Giusti, Marcelo Naiouf, Carlos García Sanchez, Guillermo Botella Juan, Manuel Prieto-Matías. "State-of-the-art in Smith-Waterman Protein Database Search". Big Data Analytics in Genomics. Ka-Chun Wong (Editor). Springer (New York), 2016. En prensa.
- [RUC16b] Rucci, Enzo. "Evaluación de rendimiento y eficiencia energética en sistemas heterogéneos para bioinformática". Tesis de Doctorado en Ciencias Informáticas (Facultad de Informática – UNLP). 2016.
- [SAN15] Sanz Victoria. "Análisis de rendimiento y optimización de algoritmos paralelos Best-First Search sobre multicore y cluster de multicore". Tesis de Doctorado de Ciencias Informáticas (Facultad de Informática – UNLP). 2015. <http://hdl.handle.net/10915/44478>.
- [SET13] High-performance Dynamic Programming on FPGAs with OpenCL. Sean Settle. 2013 IEEE High Performance Extreme Computing Conf (HPEC '13), 2013.
- [SID07] Siddh S., Pallipadi V., Mallick A. "Process Scheduling Challenges in the Era of Multicore Processors". Intel Technology Journal, Vol. 11, Issue 04, November 2007.
- [XIL15] Xilinx Inc. SDAccel Development Environment. [Online]. Disponible en <http://www.xilinx.com/products/design-tools/software-zone/sdaccel.html>

Evolución y Tendencias en Sistemas Paralelos para HPC

Marcelo Naiouf⁽¹⁾, Franco Chichizola⁽¹⁾, Laura De Giusti⁽¹⁾, Enzo Rucci⁽¹⁾⁽²⁾, Adrián Pousa⁽¹⁾, Ismael Rodríguez⁽¹⁾, Sebastián Rodríguez Eguren⁽¹⁾, Erica Montes de Oca⁽¹⁾, María José Basgall⁽¹⁾⁽²⁾, Juan Manuel Paniego⁽¹⁾, Martín Pi Puig⁽¹⁾, Emmanuel Frati⁽³⁾, Javier Balladini⁽⁴⁾, Armando De Giusti⁽¹⁾⁽²⁾

¹Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – Universidad Nacional de La Plata

²CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas

³Universidad Nacional de Chilecito

⁴Universidad Nacional del Comahue

{mnaiouf,francoch,ldgiusti,erucci,apousa,ismael,seguren,emontesdeoca,mjbasgall,jmpaniego,mpipuig,fefrati,degusti}@lidi.info.unlp.edu.ar; javier.balladini@gmail.com

Resumen

El eje de esta línea de I/D lo constituye el estudio de tendencias actuales en las áreas de arquitecturas y algoritmos paralelos. Incluye como temas centrales:

- Arquitecturas Many-core (GPU, procesadores MIC), híbridas (diferentes combinaciones de multicores y GPUs), heterogéneas, asimétricas y FPGAs.
- Cloud Computing para HPC (especialmente para aplicaciones de Big Data) y sistemas distribuidos de tiempo real (Cloud Robotics).
- Desarrollo y evaluación de algoritmos paralelos sobre nuevas arquitecturas y su evaluación de rendimiento energético y computacional.
- Empleo de contadores de hardware, en particular en toma de decisiones en tiempo de ejecución.

Palabras clave: *Sistemas Paralelos. Clusters híbridos. Arquitecturas asimétricas. GPU, MIC, FPGA. Cloud Computing y cloud robotics. Performance y eficiencia energética. Contadores de hardware.*

Contexto

Se presenta una línea de Investigación que es parte de los Proyectos 11/F018 “Arquitecturas multiprocesador en HPC: Software de Base, Métricas y Aplicaciones” y 11/F017 “Cómputo Paralelo de Altas Prestaciones. Fundamentos y Evaluación de Rendimiento en HPC. Aplicaciones a Sistemas

Inteligentes, Simulación y Tratamiento de Imágenes” del III-LIDI acreditados por el Ministerio de Educación y de proyectos específicos apoyados por organismos nacionales e internacionales. Y del proyecto “Cloud Computing y Big Data. Aplicaciones a HPC y Cloud Robotics” financiado por la Facultad de Informática de la UNLP.

En el tema hay cooperación con varias Universidades de Argentina y se está trabajando con Universidades de América Latina y Europa en proyectos financiados por CyTED, AECID y la OEI (Organización de Estados Iberoamericanos).

Por otra parte, se tiene financiamiento de Telefónica de Argentina en Becas de grado y posgrado y se ha tenido el apoyo de diferentes empresas (IBM, Microsoft, Telecom, INTEL) en la temática de Cloud Computing.

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática.

Asimismo el III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD) del MINCYT.

Introducción

Una de las áreas de creciente interés lo constituye el cómputo de altas prestaciones, en el cual el rendimiento está relacionado con dos aspectos: por un lado las arquitecturas de soporte y por el otro los algoritmos que hacen uso de las mismas.

A la aparición de arquitecturas many-core (como las GPU o los procesadores MIC), se

ha sumado el uso de FPGAs debido a su potencia de cómputo y rendimiento energético. Su combinación en sistemas HPC da lugar a plataformas híbridas con diferentes características [RUC16].

Lógicamente, esto trae aparejado una revisión de los conceptos del diseño de algoritmos paralelos (incluyendo los lenguajes mismos de programación y el software de base), así como la evaluación de las soluciones que éstos implementan. También resulta necesario investigar las estrategias de distribución de datos y procesos a fin de optimizar la performance.

Además de las evaluaciones clásicas de rendimiento prestacional como el speedup y la eficiencia, otros aspectos comienzan a ser de interés, tales como el estudio del consumo y la eficiencia energética de tales sistemas paralelos [CAS12].

Los avances en las tecnologías de virtualización y cómputo distribuido han dado origen al paradigma de Cloud Computing, que se presenta como una alternativa a los tradicionales sistemas de Clusters y Multicluster para ambientes de HPC [ROD07][BER08]. A su vez, este concepto se puede ampliar a sistemas distribuidos de tiempo real, en particular sistemas inteligentes como son los robots que pueden trabajar en paralelo utilizando su propia capacidad de procesamiento y al mismo tiempo conectándose con la potencia de un servidor en la nube (Cloud Robotics) [LOR13][GUO12][KEH15].

Una herramienta que ha comenzado a ser utilizada para la monitorización y evaluación de soluciones paralelas en tiempo de ejecución son los contadores de hardware, que (en sus diferentes variantes) permiten detectar fallas de concurrencia o estimar problemas de consumo instantáneo o decidir una migración de datos o procesos [BOR05] [SPR02].

En esta línea de I/D se trabaja sobre estos aspectos que marcan tendencias en el área.

GPUs y Cluster de GPUs

La combinación de GPUs con otras plataformas paralelas como clusters y multicores, brindan un vasto conjunto de

posibilidades de investigación en arquitecturas híbridas, a partir de diferentes combinaciones a saber:

- Cluster de máquinas uncore cada una con placa GPU, lo que permite combinar herramientas de programación paralela como MPI/CUDA.

- Máquinas multicore con más de una GPU, que combinan herramientas de programación paralela como OpenMP/CUDA o Pthread/CUDA.

- Cluster de máquinas multicore cada una con placa de GPU, lo que permite combinar OpenMP/MPI/CUDA o Pthread/MPI/CUDA.

Los desafíos que se plantean son múltiples, sobre todo en lo referido a distribución de datos y procesos en tales arquitecturas híbridas a fin de optimizar el rendimiento de las soluciones.

MIC

En forma reciente Intel brinda una alternativa a partir de los procesadores MIC (Many Integrated Core Architecture), permitiendo utilizar métodos y herramientas estándar de programación con altas prestaciones.

MIC combina varios cores Intel en un solo chip, y la programación puede realizarse usando código standard C, C++ y FORTRAN.

El mismo programa fuente escrito para MIC puede ser compilado y corrido en un Xeon, por lo que los modelos de programación familiares remueven barreras de entrenamiento y permiten enfocarse en el problema más que en la ingeniería del software.

Esta tendencia resulta promisoría, siendo el co-procesador Phi un ejemplo de esta arquitectura [JEF13].

FPGAs

Una FPGA (Field Programmable Gate Array) es una clase de acelerador basado en circuitos integrados reconfigurables. La capacidad de adaptar sus instrucciones de acuerdo a la aplicación objetivo le permite incrementar la productividad de un sistema y mejorar el rendimiento energético para ciertos tipos de aplicaciones. En la actualidad las

principales empresas fabricantes de FPGA trabajan en nuevas herramientas de programación empleando estándares familiares para HPC. Se espera que su empleo contribuya a reducir los tradicionales tiempos y costos de programación [SEA13][XIL15].

Consumo energético

Un aspecto de interés creciente en la informática actual, principalmente a partir de las plataformas con gran cantidad de procesadores, es el del consumo energético que los mismos producen.

Muchos esfuerzos están orientados a tratar el consumo como eje de I/D, como métrica de evaluación, y también a la necesidad de metodologías para medirlo.

Entre los puntos de interés pueden mencionarse:

- Caracterización energética de las instrucciones, de algoritmos complejos y de sistemas paralelos desde el punto de vista de la potencia máxima y el consumo total.
- Modelos de predicción de performance energética.
- Análisis de esquemas de distribución de procesos entre procesadores, con ajuste dinámico de la frecuencia de clock en función del consumo.
- Estudio de técnicas para reducir el consumo energético en aplicaciones de HPC de acuerdo a las arquitecturas utilizadas.
- Evaluación de eficiencia energética de diferentes algoritmos y plataformas paralelas.

Cloud Computing

Cloud Computing, proporciona grandes conjuntos de recursos físicos y lógicos (como ser infraestructura, plataformas de desarrollo, almacenamiento y/o aplicaciones), fácilmente accesibles y utilizables por medio de una interfaz de administración web, con un modelo de arquitectura “virtualizada” [SHA10][XIN12]. Estos recursos son proporcionados como servicios (“as a service”) y pueden ser dinámicamente reconfigurados para adaptarse a una carga de trabajo variable (escalabilidad), logrando una mejor utilización y evitando el sobre o sub

dimensionamiento (elasticidad) [VEL09][VAQ09].

Más allá de las potenciales características y beneficios que brinda un Cloud, de por sí atractivas, es de gran interés estudiar el despliegue de entornos de ejecución para cómputo paralelo y distribuido (Clusters Virtuales), como así también realizar I/D en la portabilidad de las aplicaciones de HPC en el Cloud [DOE11][ROD11].

Los Clusters Virtuales (VC), están conformados por VMs configuradas e interconectadas virtualmente para trabajar en forma conjunta como un recurso de cómputo único e integrado. Cada una tiene asociado un S.O., recursos de almacenamiento, protocolos de comunicación, configuraciones de red y entornos de software para ejecución de algoritmos paralelos [VAZ09][HAC11].

Por otro lado, Cloud Robotics es una de las áreas más prometedoras de la investigación informática actual en la cual se cuenta con “robots” dotados de diferentes sensores y capacidades, conectados a un Cloud vía Internet. Los temas de investigación derivados son múltiples: sensores, redes de sensores e inteligencia distribuida; robótica y sistemas colaborativos de tiempo real basados en robots; aplicaciones críticas (por ej. en ciudades inteligentes o en el ámbito industrial).

Contadores de hardware

Todos los procesadores actuales poseen un conjunto de registros especiales denominados contadores de hardware [SPR02]. Estos registros se pueden programar para contar el número de veces que ocurre un evento dentro del procesador durante la ejecución de una aplicación. Tales eventos pueden proveer información sobre diferentes aspectos de la ejecución de un programa (por ejemplo el número de instrucciones ejecutadas, la cantidad de fallos cache en L1, cuántas operaciones en punto flotante se ejecutaron, etc). Los procesadores actuales poseen una gran cantidad de eventos (más de 300) y la capacidad de usar hasta 11 registros simultáneamente [INT12]. El acceso a estos recursos de monitorización se puede llevar a

cabo usando diferentes herramientas en función del nivel de abstracción deseado.

Interesa plantear la utilización de contadores de hardware en dos líneas:

- Optimización de programas. La dependencia que tiene un programa paralelo sobre su arquitectura para ser altamente eficiente requiere comprender los motivos que penalizan su desempeño. Los contadores hardware permiten acceder a información precisa sobre aspectos específicos de la ejecución de los programas, ayudando al programador en la tarea de encontrar esos motivos y comparar con datos concretos los beneficios de los cambios que realiza [TIN14].
- Sintonización dinámica de aplicaciones. La información que proporciona la PMU (Unidad de Monitorización de Performance) de los procesadores puede ser utilizada para modificar el comportamiento o tomar decisiones en tiempo de ejecución. Esto permite construir algoritmos dinámicos de gran precisión, que se ajustan a los eventos que ocurren en el hardware [FRA14].

Líneas de Investigación, Desarrollo e Innovación

- Arquitecturas many-core (procesadores MIC y GPU) y FPGA. Análisis de este tipo de máquinas y de técnicas para desarrollar código optimizado.
- Arquitecturas híbridas (diferentes combinaciones de clusters, multicores, manycores y FPGAs). Diseño de algoritmos paralelos sobre las mismas.
- Cloud Computing para realizar HPC. Evaluación de performance en este tipo de arquitectura. Análisis del overhead por el software de administración del Cloud.
- Sistemas inteligentes de tiempo real distribuidos (Cloud Robotics).
- Migración en vivo de VCs homogéneos y heterogéneos en las arquitecturas de Cloud Computing.
- Consumo energético en las diferentes arquitecturas paralelas, en particular en relación a los algoritmos paralelos y la configuración de la arquitectura. Modelos y

predicción de performance energética en sistemas paralelos.

- Contadores de hardware. Aplicaciones en la optimización de aplicaciones paralelas.

Resultados y Objetivos

Investigación experimental a realizar

- Desarrollo y evaluación de algoritmos paralelos sobre nuevas arquitecturas. Análisis de rendimiento, eficiencia energética y costo de programación.
- Empleo experimental de contadores de hardware, orientados a la detección de fallas de concurrencia [FRA15] y a la toma de decisiones sobre la frecuencia de clock de los procesadores en función del consumo.
- Análisis del overhead introducido por el sistema gestor del Cloud en un entorno de HPC para aplicaciones científicas de Big Data. Comparar el rendimiento entre Cloud y Cluster Computing.
- Analizar y comparar las técnicas de migración en vivo de VMs, con el fin de implementar migraciones de VCs homogéneos y heterogéneos en Cloud Computing.
- Realizar el desarrollo de nuevos planificadores de tareas para multicores asimétricos sobre diferentes sistemas operativos con el objetivo de maximizar el rendimiento y minimizar el consumo de energía [SAE14][SAE15].
- Optimización de algoritmos paralelos para controlar el comportamiento de múltiples robots que trabajan colaborativamente, considerando la distribución de su capacidad de procesamiento “local” y la coordinación con la potencia de cómputo y capacidad de almacenamiento (datos y conocimiento) de un Cloud.

Resultados obtenidos

- En la línea de planificación de procesos sobre multicores asimétricos se evaluaron nuevos algoritmos de planificación sobre los kernel Solaris y Linux, se analizaron las mejoras en rendimiento y consumo energético para un conjunto diverso de cargas de trabajo [SAE14][SAE15].

- Utilización combinada de cluster de multicores y cluster de GPUs [MON14][POU15].

- Se ha finalizado una tesis de doctorado focalizada en evaluación de rendimiento y eficiencia energética de diferentes plataformas heterogéneas, en la cual se desarrollaron diferentes soluciones paralelas para un problema del área bioinformática con alta demanda computacional [RUC16]. Se analizaron sus rendimientos y el consumo energético y costo de programación asociados.

Organización de Eventos

En el año 2015 se ha organizado la III Jornada de Cloud Computing y Big Data (JCC&BD 2015) en Argentina, con participación de especialistas académicos del país y del exterior y de empresas con experiencia en Cloud Computing. En junio de 2016 se organizará la IV Jornadas de Cloud Computing y Big Data (JCC&BD 2016).

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D el último año se concluyeron: 2 tesis doctorales, 2 trabajos de Especialización y 1 tesina de grado. Al mismo tiempo se encuentran en curso 4 tesis de Doctorado en Ciencias Informáticas.

Además, se participa en el dictado de las carreras de Doctorado en Ciencias Informáticas, y Magíster y Especialización en Cómputo de Altas Prestaciones de la Facultad de Informática de la UNLP (todas acreditadas A por la CONEAU), por lo que potencialmente pueden generarse nuevas Tesis de Doctorado y Maestría y Trabajos Finales de Especialización.

Existe cooperación con grupos de otras Universidades del país y del exterior, y hay tesis de diferentes Universidades realizando su Tesis con el equipo del proyecto.

Respecto a las carreras de grado, se dictan por parte de integrantes de la línea de investigación dos materias directamente relacionadas con los temas de la misma: "Taller de Programación sobre GPUs" y

"Cloud Computing. Aplicaciones en Big Data"

Referencias

[BER08] Bertogna, M., Grosclaude, E., Naiouf, M., De Giusti, A., Luque, E.: "Dynamic on Demand Virtual Clusters in Grids". In: 3rd Workshop on Virtualization in High-Performance Cluster and Grid Computing (VHPC 08). España. (2008).

[BOR05] S. Y. Borkar, P. Dubey, K. C. Kahn, D. J. Kuck, H. Mulder, S. S. Pawlowski, y J. R. Rattner: "Platform 2015: Intel® Processor and Platform Evolution for the Next Decade". In: Intel Corporation, White Paper, 2005.

[CAS12] Casanova B., Balladini J., De Giusti A., Suppi R., Rexachs D., Luque E.: "Mejora de la eficiencia energética en sistemas de computación de altas prestaciones". In: XII Workshop de Procesamiento Distribuido y Paralelo. CACIC 2012. ISBN: 978987-1648-34-4. Pág. 377-386. Argentina (2012).

[DOE11] Doelitzcher, F., Held, M., Sulistio, A., Reich, C. Viteras: "Virtual Cluster as a Service". In: 3rd IEEE International Conference on Cloud Computing Technology and Science. Grecia (2011).

[FED09] Alexandra Fedorova, Juan Carlos Saez, Daniel Shelepov, Manuel Prieto: "Maximizing Power Efficiency with Asymmetric Multicore Systems". In: Communications of the ACM, Vol. 52 (12), pp 48-57. 2009.

[FRA14] F. E. Frati, K. Olcoz Herrero, L. Piñuel Moreno, M. R. Naiouf, A. De Giusti: "Modelo para sintonización dinámica de aplicaciones guiada por eventos de hardware". In: XX Congreso Argentino de Ciencias de la Computación. 2014.

[FRA15] Frati E: "Software para arquitecturas basadas en procesadores de múltiples núcleos. Detección Automática de errores de concurrencia". Tesis Doctoral, Universidad Nacional de La Plata. 2015.

[GUO12] Guoqiang Hu, Wee Peng Tay, Yonggang Wen: "Cloud robotics: architecture, challenges and applications". In: Network, IEEE, vol.26, no.3, pp.21-28. 2012.

[HAC11] Hacker, T., Mahadik, K.: "Flexible Resource Allocation for Reliable Virtual Cluster Computing." In: Supercomputing Proceedings (SC11). USA (2011).

[INT12] Intel Corporation: "Intel® 64 and IA-32 Architectures Software Developer's Manual". Manual 253669-043US, 2012. Recuperado a partir de <http://www.intel.com/content/www/us/en/processors/architectures-software-developer-manuals.html>

[JEF13] Jeffers, James; Reinders, James: "Intel Xeon Phi Coprocessor High Performance Programming", Morgan Kaufmann, 2013.

[KEH15] Kehoe B., Patil S., Abbeel P., Goldberg K.: "A Survey of Research on Cloud Robotics and Automation". In: IEEE Transactions on Automation Science and Engineering (T-ASE): Special Issue on Cloud Robotics and Automation. Vol. 12, no. 2. 2015.

- [LOR13] Lorencik D., Sincak P.: "Cloud robotics: Current trends and possible use as a service". In: Applied Machine Intelligence and Informatics (SAMI), 2013 IEEE 11th International Symposium on , vol., no., pp.85-88. 2013.
- [POU15] Adrian Pousa, Victoria Sanz, Armando De Giusti: "Comparación de rendimiento de algoritmos de cómputo intensivo y de acceso intensivo a memoria sobre arquitecturas multicore. Aplicación al algoritmo de criptografía AES". XXI Congreso Argentino de Ciencias de la Computación. Argentina (2015).
- [ROD07] Rodriguez, I. P., Pousa, A., Pettoruti, J.E., Chichizola, F., Naiouf, M., De Giusti, L., De Giusti, A.: "Estudio del overhead en la migración de algoritmos paralelos de cluster y multicluster a GRID". In: Proceedings del XIII Congreso Argentino de Ciencias de la Computación. Argentina (2007).
- [ROD11] Rodriguez, I., Pettoruti, J.E., Chichizola, F., De Giusti, A.: "Despliegue de un Cloud Privado para entornos de cómputo científico". In: Proceedings del XVII Congreso Argentino de Ciencias de la Computación. Argentina (2011).
- [RUC16] Rucci, Enzo: "Evaluación de rendimiento y eficiencia energética en sistemas heterogéneos para bioinformática". Tesis de Doctorado en Ciencias Informáticas (Facultad de Informática – UNLP). 2016.
- [SAE14] Juan Carlos Saez, Adrian Pousa, Fernando Castro, Daniel Chaver, Manuel Prieto: "Exploring the Throughput-Fairness Trade-off on Asymmetric Multicore Systems". In: Euro-Par 2014: Parallel Processing Workshops - 2nd Workshop on Runtime and Operating Systems for the Many-core Era (ROME 2014). 2014.
- [SAE15] Juan Carlos Saez, Adrian Pousa, Daniel Chaver, Fernando Castro, Manuel Prieto Matias: "ACFS: A Completely Fair Scheduler for Asymmetric Single-ISA Multicore Systems". In: ACM SAC 2015 (The 30TH ACM/SIGAPP Symposium on applied computing). 2015.
- [SEA13] Sean Settle: "High-performance Dynamic Programming on FPGAs with OpenCL". In: IEEE High Performance Extreme Computing Conference. 2013.
- [SHA10] Shafer, J.: "I/O virtualization bottlenecks in cloud computing today". In: Proceedings of the 2nd conference on I/O virtualization (VIOV10). USA (2010).
- [SPR02] B. Sprunt: "The basics of performance-monitoring hardware". In: IEEE Micro, vol. 22, nro 4, pp. 64- 71. 2002.
- [TIN14] F. G. Tinetti, S. M. Martin, F. E. Frati, M. Méndez: "Optimization and Parallelization experiences using hardware performance counters". In: Proceedings of the 4th International Supercomputing Conference. México (2013).
- [VAQ09] Vaquero, L. M., Rodero-Merino, L., Caceres, J., Lindner, M.: "A Break in the Clouds: Towards a Cloud Definition". In: ACM SIGCOMM Computer Communication Review, Volume 39, Issue 1, pp. 50--55. USA. (2009).
- [VAZ09] Vázquez Blanco, B., Huedo, E., Montero, R. S., Llorente, I. M.: "Elastic Management of Cluster-based Services in the Cloud". In: Proceedings pp.19-24, ACM Digital Library. ISBN 978-1-60558-564-2. 2009.
- [VEL09] Velte, A.T., Velte, T.J., Elsenpeter, R.: "Cloud Computing: A Practical Approach". McGraw Hill Professional. 2009.
- [XIL15] Xilinx Inc. "SDAccel Development Environment". [Online]. Disponible en <http://www.xilinx.com/products/design-tools/software-zone/sdaccel.html>
- [XIN12] Xing, Y., Zhan, Y.: "Virtualization and Cloud Computing". In: Proceedings pp.305-312, Springer Link. ISBN 978-3-642-27323-0. (2012). Morgan Kaufmann. 2013.

Arquitecturas Multiprocesador en HPC: Software, Métricas y Aplicaciones.

De Giusti Armando^{1,2}, Tinetti Fernando^{1,3}, Naiouf Marcelo¹, Chichizola Franco¹, De Giusti Laura¹, Villagarcía Horacio^{1,3}, Montezanti Diego¹, Encinas Diego¹, Pousa Adrián¹, Rodríguez Ismael¹, Eguren Sebastián¹, Iglesias Luciano¹, Paniego Juan Manuel⁽¹⁾, Pi Puig Martín⁽¹⁾, Dell'Oso Matías⁽¹⁾, Mendez Mariano⁴

¹Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – Universidad Nacional de La Plata

²CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas

³CIC - Comisión de Investigaciones de la Provincia de Buenos Aires

⁴Facultad de Ingeniería - Universidad de Buenos Aires

{degiusti, fernando, mnaiouf, francoch, ldgiusti, hvw, dmontezanti, dencinas, apousa, ismael, seguren, li, mmendez, jmpaniego, mpipuig, mdelloso}@lidi.info.unlp.edu.ar, marianomendez@gmail.com

Con la colaboración en la dirección de Tesis de Posgrado de la Universidad Autónoma de Barcelona (España) y la Universidad Complutense de Madrid (España).

Resumen

Caracterizar las arquitecturas multiprocesador distribuidas enfocadas especialmente a cluster y cloud computing, con énfasis en las que utilizan procesadores de múltiples núcleos (multicores, GPUs y Xeon Phi), con el objetivo de modelizarlas, estudiar su escalabilidad, analizar y predecir performance de aplicaciones paralelas, estudiar el consumo energético y su impacto en la performance así como desarrollar esquemas para detección y tolerancia a fallas en las mismas.

Profundizar el estudio de arquitecturas basadas en GPUs y su comparación con clusters de multicores, así como el empleo combinado de GPUs y multicores en computadoras de alta performance. Iniciar investigación experimental con arquitecturas paralelas basadas en FPGAs. En particular estudiar performance en Clusters “híbridos”.

Analizar y desarrollar software de base para clusters, tratando de optimizar el rendimiento.

Investigar arquitecturas multicore asimétricas, desarrollar algoritmos de planificación en el software de sistema operativo para permitir la optimización del rendimiento y consumo energético en aplicaciones de propósito general.

Estudiar clases de aplicaciones inteligentes en tiempo real, en particular el trabajo colaborativo de robots conectados a un cloud.

Es de hacer notar que este proyecto se coordina con otros proyectos en curso en el III-LIDI, relacionados con Algoritmos Paralelos, Sistemas Distribuidos y Sistemas de Tiempo Real.

Palabras claves: *Sistemas Paralelos. Multicore. GPU. FPGAs, Cluster y Cloud Computing. Cluster híbridos. Performance y eficiencia energética. Tolerancia a fallas en Sistemas paralelos. Modelos de programación de arquitecturas paralelas. Planificación. Scheduling. Cloud Robotics.*

Contexto

Esta línea de Investigación está dentro del proyecto 11/F018: “Arquitecturas multiprocesador en HPC: Software de base, Métricas y Aplicaciones acreditado por el Ministerio de Educación y de proyectos específicos apoyados por organismos nacionales e internacionales.

El III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD) del MINCYT y en esta línea de I/D hay cooperación con varias Universidades de Argentina y se está trabajando con Universidades de América Latina y Europa en proyectos financiados por CyTED, AECID y la OEI (Organización de Estados Iberoamericanos).

En la Facultad de Informática de la UNLP (a partir del equipo del proyecto) se han incorporado asignaturas optativas en la curricula de grado de las carreras de Licenciatura en Informática, Licenciatura en Sistemas e Ingeniería en Computación relacionadas con Cloud Computing, Big Data y Programación sobre GPGPUs. Además, la Facultad aprobó y financia el proyecto “Cloud Computing y Big Data. Aplicaciones a HPC y Cloud Robotics”.

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática, así como el desarrollo de la Maestría y Especialización en Computación de Altas Prestaciones, acreditadas por CONEAU.

Por último, se tiene financiamiento de Telefónica de Argentina en Becas de grado y posgrado y se ha tenido el apoyo de diferentes empresas (IBM, Microsoft, Telecom, INTEL, AMAZON AWS) en las temáticas de Cloud Computing y Big Data.

Introducción

La investigación en Paralelismo (a partir de arquitecturas multiprocesador distribuidas o

concentradas en supercomputadoras) es una de las líneas de mayor desarrollo en la Ciencia Informática actual [GRA03]. La utilización de clusters, multiclusters, grids y clouds, soportadas por redes de diferentes características y topologías se ha generalizado, tanto para el desarrollo de algoritmos paralelos orientados a HPC como para el manejo de aplicaciones distribuidas y/o servicios WEB concurrentes [GRA03][MCC12][DON03][BEN06].

El cambio tecnológico, fundamentalmente a partir de los procesadores multicore, ha impuesto la necesidad de investigar en paradigmas "híbridos", en los cuales coexisten esquemas de memoria compartida con mensajes [LEI12][POU11]. Asimismo la utilización de aceleradores (GPU, FPGA, Xeon Phi) presenta una alternativa para alcanzar un alto speedup en determinadas clases de aplicaciones [MON12][KIN09][SIN12][LIN11]. Debe notarse que el modelo de programación orientado a estas arquitecturas cambia sensiblemente y la optimización de código paralelo requiere nuevos recursos.

Por otro lado, recientemente han aparecido los procesadores multicore asimétricos (AMPs) que integran en un mismo chip diversos tipos de cores con distintas características (frecuencia, microarquitectura o consumo), pero con el mismo repertorio de instrucciones. Investigaciones previas han demostrado que los AMPs ofrecen un mayor rendimiento por watt y unidad de área que los multicores simétricos [SAE10][ANN12]. Además se ha demostrado que la integración de sólo dos tipos de core (rápidos y lentos) en el chip permite una utilización muy eficiente del área del procesador y simplifica considerablemente el diseño.

Es importante en este contexto re-analizar el concepto de *eficiencia* incluyendo tanto al aspecto computacional como el energético y considerar el impacto del consumo sobre arquitecturas con miles de procesadores que trabajan concurrentemente, este concepto extendido de eficiencia conduce a un estudio de nuevos lenguajes, paradigmas y herramientas [RUC16][BAL12].

Las arquitecturas tipo "*Cloud*" se presentan como una evolución natural del concepto de *Clusters* y *Grids*, integrando grandes conjuntos de recursos virtuales (hardware, plataformas de desarrollo y/o servicios), fácilmente accesibles y utilizables por usuarios distribuidos, vía WEB [MIL08][VAQ09][FOS10]. Estos recursos pueden ser dinámicamente reconfigurados para adaptarse a una carga variable, permitiendo optimizar su uso. Al enfocarnos en Cloud Computing aparecen problemas clásicos de la Ciencia Informática, extendidos para este nuevo modelo de arquitectura: planificación, virtualización, asignación dinámica de recursos, migración de datos y procesos [ARD09][VAZ09]. En el proyecto se ha

abierto una línea específicamente dedicada a Cloud Computing, incluyendo los temas de configuración y administración eficiente de Cloud, así como las aplicaciones sobre Cloud incluyendo las de "big data", y aquellas que requieren centralizar el accionar de "robots" distribuidos en tiempo real (Cloud Robotics) [MAY13].

Las aplicaciones científicas con un uso intensivo de datos utilizan software de E/S paralelo para acceder a archivos. Contar con herramientas que permitan predecir el comportamiento de este tipo de aplicaciones en HPC es de gran utilidad para los desarrolladores de aplicaciones paralelas. Por otro lado, el modelado basado en agentes y simulación (Agent-Based Modeling and Simulation, ABMS) [MAC06] ha sido utilizado para modelar problemas y sistemas complejos en diversas áreas de la ciencia. También es importante focalizar la investigación en los problemas de detección y tolerancia a fallos en arquitecturas paralelas, tratando de minimizar el overhead temporal y aprovechar la redundancia de hardware (nivel de ocupación de los núcleos) que caracteriza estas arquitecturas [GOL09][FIA11]. El manejo de fallos es una preocupación creciente en HPC; en el futuro, se esperan mayores tasas de errores, intervalos de detección más largos y fallos silenciosos, que tienen la capacidad de corromper los resultados de las aplicaciones. Debido a ello, sus impactos se vuelven relevantes, en especial cuando se consideran aplicaciones científicas de gran duración, debido al alto costo de relanzar la ejecución desde el comienzo en caso de resultados incorrectos. Se proyecta que, en sistemas de exaescala, los errores ocurran varias veces al día y se propaguen para generar desde caídas de procesos hasta corrupciones de resultados, con fallos no detectados en aplicaciones que siguen operando. Por ello, se han desarrollado herramientas de detección de fallos transitorios [MON12][MON14][MON15] basadas en replicación de software, detectando divergencias en las comunicaciones entre réplicas y evitando que la corrupción se propague a otros procesos y restringiendo la latencia de detección. De esta forma se permiten obtener ejecuciones fiables con resultados correctos o conducir al sistema a una parada segura.

Interesan también los problemas que significan integración de redes de sensores con modelos del mundo real (por ej. modelos meteorológicos, hídricos o de terreno) para prevención de emergencias [SIN06][GAU16]. En esta línea, el eje del proyecto sigue estando en la problemática del paralelismo combinado con sistemas de tiempo real, pudiendo contribuir a proyectos multidisciplinarios, en particular por temas de emergencias hídricas, exploración de recursos naturales y temas de atención sanitaria y evacuación de edificios en situaciones de emergencia.

Líneas de Investigación, Desarrollo e Innovación

- Estudio y caracterización de arquitecturas paralelas: clusters, grids, clouds, aceleradores (GPU, FPGA, Xeon Phi), híbridos.
- Desarrollo de algoritmos de planificación de procesos orientado a procesadores asimétricos es una línea de interés por la posibilidad de especializar algunos núcleos a funciones específicas, optimizando el rendimiento general. Investigación en los diferentes niveles: sistema operativo, compiladores, técnicas de programación.
- Desarrollo de aplicaciones concretas (numéricas y no numéricas) sobre diferentes arquitecturas paralelas, y el estudio de paradigmas y patrones de programación. Técnicas de optimización de aplicaciones paralelas.
- Desarrollo de técnicas de tolerancia a fallas en sistemas paralelos y distribuidos, lo cual supone un avance en el estudio de las formas de aprovechar la redundancia de dichas arquitecturas que no resulten eficientemente utilizadas.
- Desarrollo de herramientas para la transformación de código heredado, buscando su optimización sobre arquitecturas paralelas.
- Integración de métricas de rendimiento computacional y energético. Predicción de performance de aplicaciones paralelas.
- Cloud Computing. Software de base. Desarrollo de aplicaciones de HPC (principalmente de big data).
- Sistemas inteligente distribuidos de tiempo real conectados al Cloud (Cloud Robotics).

Resultados y Objetivos

- Estudiar modelos complejos, que integren redes de sensores en tiempo real y cómputo paralelo. Estrategias de predicción de catástrofes (inundaciones, incendios por ejemplo) se basan en estos modelos con alta capacidad de procesamiento y monitoreo de señales en tiempo real [GAU16].
- Se han desarrollado diferentes aplicaciones adaptadas para diferentes arquitecturas "híbridas" (que combinan multicore y aceleradores), y analizado/comparado el rendimiento obtenido [RUC16][POU15][MON12].
- Analizar y adaptar los patrones de programación eficiente de algoritmos híbridos que exploten simultáneamente el paradigma de mensajes y el de memoria compartida.[POU15][RUC16][MON12].
- Se está trabajando en técnicas de recuperación desde fallas transitorias a partir de múltiples puntos de verificación (checkpoints), que sirvan para garantizar la correcta finalización de

aplicaciones científicas sobre sistemas de HPC, que resultan afectadas por la ocurrencia de fallas externas y aleatorias [MON14][MON15]. Para ello deben realizarse múltiples checkpoints incrementales distribuidos (sin sobrescritura) y desarrollar técnicas de selección de versión en función de la presencia de un fallo latente no detectado previamente, integrando esta solución con las herramientas de detección desarrolladas previamente.

- Realizar el desarrollo de nuevos planificadores de tareas para multicore asimétricos sobre diferentes sistemas operativos con el objetivo de maximizar el rendimiento y minimizar el consumo de energía [SAE15][POU15].
- Desarrollo de un modelo de la Entrada/Salida en HPC por medio de ABMS (Agent-Based Modeling and Simulation) que permita predecir cómo cambios realizados en los diferentes componentes del modelo afectan a la funcionalidad y el rendimiento del sistema [ENC15].
- Desarrollo de aplicaciones vinculadas con "Big Data", especialmente para resolver en Cloud Computing [BAS15]. (en relación con los otros proyectos del III-LIDI).
- Optimización de algoritmos paralelos para controlar el comportamiento de múltiples robots que trabajan colaborativamente, considerando la distribución de su capacidad de procesamiento "local" y la coordinación con la potencia de cómputo y capacidad de almacenamiento (datos y conocimiento) de un Cloud.
- Actualización y modernización de código fuente de Sistemas Heredados (Legacy Systems) de Cómputo Científico a través de la aplicación de un proceso de desarrollo iterativo e incremental dirigido por transformaciones de código fuente, apoyado fuertemente en las herramientas de desarrollo. Dichas transformaciones se implementan para ser aplicadas automáticamente en un entorno integrado de desarrollo [MEN14].
- Trabajar en la implementación de transformaciones que ayuden a la paralelización del código fuente, así como también en herramientas de análisis estático de Código fuente [TIN13].
- Adaptar las técnicas de scheduling y mapeo de procesos a procesadores de acuerdo a los objetivos actuales (en particular los relacionados con el consumo), considerando la migración dinámica de datos y procesos en función de rendimiento y consumo [GRA03][DEG10]. Se debe incluir la utilización de los registros de hardware de los procesadores para la toma de diferentes decisiones en tiempo de ejecución.
- Analizar y comparar las técnicas de migración en vivo de VMs, con el fin de implementar migraciones de VCs homogéneos y heterogéneos en Cloud Computing.

Formación de Recursos Humanos

En cooperación con Universidades iberoamericanas se ha implementado la Maestría en Cómputo de Altas Prestaciones y se continúa dictando la Especialización en Cómputo de altas Prestaciones y Tecnología GRID. Asimismo se tiene un importante número de doctorandos (del país y del exterior) realizando el Doctorado en Ciencias Informáticas de la UNLP.

Se han organizado las III Jornadas de Cloud Computing & Big Data (JCC&BD), y se realizarán en junio de 2016 las IV JCC&BD, integrando una Escuela con cursos de Posgrado relacionados con la temática.

Existe cooperación a nivel nacional e internacional y dentro de la temática del proyecto se espera alcanzar 8 Tesis de Doctorado y 6 Tesis de Maestría en los próximos 4 años, en el país. Al menos tener 3 Doctorandos en el exterior o mixtos en el mismo período.

En 2015 se aprobaron 5 Tesis Doctorales [FRA15][WOL15][MIC15][RUC16][GAU15] y otra está entregada para evaluación. También se aprobaron 6 trabajos de Especialista, 1 Tesis de Magister y 2 Tesinas de grado.

Referencias

[ANN12] Annamalai A., Rodrigues R., Koren I., Kundu S., "Dynamic Thread Scheduling in Asymmetric Multicores to Maximize Performance-per-Watt," 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum, pp. 964-971, 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum, 2012.

[ARD09] Ardissono L., Goy A., Petrone G., Segnan M. "From Service Clouds to User-centric Personal Clouds". 2009 IEEE Second International Conference on Cloud Computing.

[BAL12] Balladini J., Rucci E., De Giusti A., Naiouf M., Suppi R., Rexachs D., Luque E. "Power Characterisation of Shared-Memory HPC Systems". XII Workshop de Procesamiento Distribuido y Paralelo. CACIC 2012. ISBN: 978987-1648-34-4. Pág. 316-326. Bahía Blanca, Buenos Aires, Argentina, Octubre 2012.

[BAS15] Basgall M. J., Aquino G., Lanzarini L., Naiouf M. "Un Enfoque Dinámico para la Detección de Relaciones entre Tópicos en Textos provenientes de Redes Sociales". III Jornadas de Cloud Computing y Big Data (JCC&BD). Facultad de Informática. UNLP. 2015.

[BEN06] Ben-Ari, M. "Principles of Concurrent and Distributed Programming, 2/E". Addison-Wesley, 2006.

[DEG10] De Giusti L., Chichizola F., Naiouf M., De Giusti A.E., Luque E. "Automatic Mapping Tasks to Cores - Evaluating AMTHA Algorithm in Multicore Architectures". IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 2, No 1, March 2010. ISSN (Online): 1694-0784. ISSN (Print): 1694-0814. Págs. 1-6.

[DON03] Dongarra J., Foster I., Fox G., Gropp W., Kennedy K., Torczon L., White A. "The Sourcebook of Parallel

Computing". Morgan Kauffman Publishers. Elsevier Science, 2003.

[ENC15] D. Encinas et al., Modeling I/O System in HPC: An ABMS Approach. The Seventh International Conference on Advances in System Simulation (SIMUL), ISBN: 978-1-61208-442-8, 2015

[FIA11] Fialho L. "Fault Tolerance configuration for uncoordinated checkpoints". Ph.D. Thesis, Universidad Autónoma de Barcelona, Barcelona, España, Julio 2011.

[FRA15] Frati E. "Software para arquitecturas basadas en procesadores de múltiples núcleos. Detección Automática de errores de concurrencia" Tesis Doctoral, Universidad Nacional de La Plata, Marzo 2015.

[GAU15] Adriana Gaudiani. "Simulación y optimización como metodología para mejorar la calidad de la predicción en un entorno de simulación hidrográfica". Tesis de Doctorado en Ciencias Informáticas (Facultad de Informática – UNLP). 2015.

[GAU16] Adriana Gaudiani, Emilio Luque, Pablo García, Mariano Re, Marcelo Naiouf, Armando De Giusti. "How a Computational Method Can Help to Improve the Quality of River Flood Prediction by Simulation". Advances and New Trends in Environmental and Energy Informatics (part V). ISBN 978-3-319-23455-7. Pp337-351. 2016.

[GOL09] Golander A., Weiss S., Ronen R. "Synchronizing Redundant Cores in a Dynamic DMR Multicore Architecture". IEEE Transactions on Circuits and Systems II: Express Briefs Volume 56, Issue 6, 474-478. 2009.

[GRA03] Grama A, Gupta A, Karypis G, Kumar V. "Introduction to parallel computing". Second Edition. Pearson Addison Wesley, 2003.

[KIN09] Kindratenko, V.V.; Enos, J.J.; Guochun Shi; Showerman, M.T.; Arnold, G.W.; Stone, J.E.; Phillips, J.C.; Wen-Mei Hwu, "GPU clusters for high-performance computing," Cluster Computing and Workshops, 2009. CLUSTER '09. IEEE International Conference on , vol., no., pp.1,8, Aug. 31 2009-Sept. 4 2009

[LEI12] Leibovich F., Chichizola F., De Giusti L., Naiouf M., Tirado Fernández F., De Giusti A. "Programación híbrida en clusters de multicore. Análisis del impacto de la jerarquía de memoria". XII Workshop de Procesamiento Distribuido y Paralelo. CACIC2012. ISBN: 978987-1648-34-4. Pág. 306-315. 2012.

[LIN11] Lingyan Wang, Miaoqing Huang, and Tarek El-Ghazawi. "Towards efficient GPU sharing on multicore processors". In Proceedings of the second international workshop on Performance modeling, benchmarking and simulation of high performance computing systems (PMBS '11). ACM, New York, NY, USA, 23-24. 2011.

[MAC06] C. Macal, M. North, Tutorial on agent-based modeling and simulation part 2: how to model with agents, in: Proceedings of the Winter Simulation Conference, 2006.

[MAY13] Mayer-Schönberger V., Cukier K., "Big Data: A revolution that will transform how we live, work and think" Houghton Mifflin Harcourt 2013.

[MEN14] "Integrating Software Metrics for Fortran Legacy into an IDE". Mariano Mendez, Fernando G. Tinetti. XI Workshop Ingeniería de Software – CACIC 2014. ISBN 978-987-3806-05-6. Pág. 771-780. San Justo, Buenos Aires, Argentina. Octubre 2014.

[MCC12] McCool M., Robison A., Reinders J. "Structured Parallel Programming" Elsevier-Morgan Kaufmann, 2012.

- [MIC15] Micolini O. "Arquitectura asimétrica multicore con procesador de Petri" Tesis Doctoral, Universidad Nacional de La Plata, Marzo 2015.
- [MIL08] Miller M. "Cloud computing: web-based applications that change the way you work and collaborate online". Que Publishing. USA 2008.
- [MON12] Diego Montezanti, Dolores Rexachs, Enzo Rucci, Emilio Luque, Marcelo Naiouf, Armando De Giusti. "Caracterización de una estrategia de detección de fallos transitorios en HPC". XXI CACIC 2015 (Congreso de Ciencias de la Computación) - XV Workshop de Procesamiento Paralelo y Distribuido (WPDP). ISBN: 978-987-3724-37-4.
- [MON12] Diego Montezanti, Fernando Emmanuel Frati, Dolores Rexachs, Emilio Luque, Marcelo Naiouf, Armando De Giusti. "SMCV: a Methodology for Detecting Transient Faults in Multicore Clusters". CLEI Electronic Journal, Volumen 15, Número 3 (Diciembre 2012). Páginas: 49 a 59. ISSN 0717- 5000.
- [MON12] Montes de Oca E., De Giusti L., De Giusti A., Naiouf M. "Comparación del uso de GPU y cluster de multicore en problemas con alta demanda computacional". XII Workshop de Procesamiento Distribuido y Paralelo. CACIC2012. ISBN: 978987-1648-34-4. Pág. 267-275. Bahía Blanca, Buenos Aires, Argentina, Octubre 2012.
- [MON14] Diego Montezanti, Enzo Rucci, Dolores Rexachs, Emilio Luque, Marcelo Naiouf, Armando De Giusti. "A tool for detecting transient faults in execution of parallel scientific applications on multicore clusters". Journal of Computer Science & Technology (JCS&T), Volumen 14, Número 1 (Abril 2014). Páginas: 32 a 38. ISSN: 1666-6038.
- [FOS10] Foster I. "There's Grid in them thar Clouds". 2 de Enero, 2008.<http://ianfoster.typepad.com/blog/2008/01/theres-grid-in.html>. Noviembre, 2010.
- [POU11] Pousa A., Sanz V., De Giusti A. "Análisis de rendimiento de un algoritmo de criptografía simétrica sobre arquitecturas multicore", Proceedings del XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011), La Plata (Argentina), 2011, ISBN: 978-950-34-0756-1. Págs: 231-240.
- [POU15] Adrian Pousa, Victoria Sanz, Armando De Giusti. Comparación de rendimiento de algoritmos de cómputo intensivo y de acceso intensivo a memoria sobre arquitecturas multicore. Aplicación al algoritmo de criptografía AES. CACIC 2015 (XXI Congreso Argentino de Ciencias de la Computación). 2015.
- [RUC16] Rucci, Enzo. "Evaluación de rendimiento y eficiencia energética en sistemas heterogéneos para bioinformática". Tesis de Doctorado en Ciencias Informáticas (Facultad de Informática – UNLP). 2016.
- [SAE10] Sáenz J. C. "Planificación de Procesos en Sistemas Multicore Asimétricos". Ph.D. Thesis, Universidad Complutense de Madrid, Madrid, España, Diciembre 2010.
- [SAE15] Juan Carlos Saez, Adrian Pousa, Daniel Chaver, Fernando Castro, Manuel Prieto Matias. ACFS: A Completely Fair Scheduler for Asymmetric Single-ISA Multicore Systems. ACM SAC 2015 (The 30TH ACM/SIGAPP Symposium on applied computing). 2015.
- [SIN06] Singh B. D. "Influence of Digital Elevation Models Derived from Remote Sensing on SpatioTemporal Modelling of Hydrologic and Erosion Processes". Thesis for Degree of Doctor of Philosophy. pp 370. Cranfield University Siloe England. 2006.
- [SIN12] Sinha, R.; Prakash, A.; Patel, H.D., "Parallel simulation of mixed-abstraction SystemC models on GPUs and multicore CPUs," Design Automation Conference (ASP-DAC), 2012 17th Asia and South Pacific, pp.455,460, Jan. 30 2012-Feb. 2 2012.
- [TIN13] "Restructuring Fortran legacy applications for parallel computing in multiprocessors". Fernando G. Tinetti, Mariano Méndez, Armando De Giusti. The Journal of Supercomputing, May 2013, Volume 64, Issue 2, ISSN 0920-8542, DOI 10.1007/s11227-012-0863-x, pp 638-659.
- [VAQ09] Vaquero L.M. et al. "A Break in the Clouds: Towards a Cloud Definition". ACM SIGCOMM Computer Communication Review, vol. 39, num. 1, páginas 50-55, ISSN 0146-4833. Enero 2009.
- [VAZ09] Vázquez Blanco C., Huedo E., Montero R. S., Llorente I. M. "Elastic Management of Cluster-based Services in the Cloud". Proceedings pp. 19-24, ACM Digital Library 2009. ISBN 978-1-60558-564-2.
- [WOL15] Wolfmann G. "Parallel Execution Model based on Petri Nets" Tesis Doctoral, Universidad Nacional de La Plata, Febrero 2015.

Computación de Altas Prestaciones: Problemáticas y Aplicaciones

Javier Balladini, Marina Morán, Claudia Rozas, Emmanuel Frati¹, Armando De Giusti², Remo Suppi³, Dolores Rexachs³, Emilio Luque³

Facultad de Informática, Universidad Nacional del Comahue
Buenos Aires 1400, Neuquén Capital, Tel. +54 299 4490300
{javier.balladini, marina, claudia.rozas}@fi.uncoma.edu.ar

¹ Universidad Nacional de Chilecito

La Rioja
fefrati@undec.edu.ar

² Instituto de Investigación en Informática LIDI, Universidad Nacional de La Plata

50 y 120 - 2do Piso, La Plata, Tel. +54 221 4227707

degiusti@lidi.info.unlp.edu.ar

³ Departamento de Arquitectura de Computadores y Sistemas Operativos, Universidad Autónoma de Barcelona

Campus UAB, Edifici Q, 08193 Bellaterra (Barcelona), España. Tel: +34 93 581 1990

{remo.suppi, dolores.rexachs, emilio.luque}@uab.es

Resumen

Nuestra investigación está centrada en dos líneas. Por un lado, el estudio del consumo energético de los sistemas de Cómputo de Altas Prestaciones (HPC, de High Performance Computing) cuya alta demanda energética tiene serias consecuencias financieras, medioambientales, y en muchos casos también sociales. El aumento de la eficiencia energética de los sistemas de HPC no solo proviene de las nuevas arquitecturas hardware, también está involucrado el software, quien debe gestionar y configurar el sistema para mantener un determinado equilibrio entre tiempo de ejecución, eficiencia energética y productividad. Es en este aspecto, el software, en el que estamos trabajando. Por otro lado, nos interesa el estudio de aplicaciones que utilizan sistemas de cómputo de altas prestaciones orientadas a su uso en centros de salud. En particular, nos enfocamos en sistemas de monitoreo inteligente para Unidades de Vigilancia Intermedia y Unidades de Terapia Intensiva.

Estas líneas de investigación se desarrollan en colaboración entre tres Universidades (dos nacionales y una extranjera) y un Hospital.

Palabras claves: computación ecológica, eficiencia energética, computación de altas prestaciones, aplicaciones para la salud.

1. Contexto

La investigación aquí presentada surge como una colaboración entre la Facultad de Informática de la Universidad Nacional del Comahue (UNCOMA), el Instituto de Investigación en Informática LIDI de la Universidad Nacional de La Plata (UNLP), y el grupo de investigación HPC4EAS (High Performance Computing for Efficient Applications and Simulation) de la Universidad Autónoma de Barcelona (España). La colaboración se centra actualmente en dos temáticas. Una de ellas se centra en la problemática del consumo energético de los sistemas de cómputo paralelo al ejecutar aplicaciones científicas computacionalmente complejas. La otra temática está relacionada a aplicaciones para instituciones de la salud que requieren cómputo de altas prestaciones. En este último caso, se está trabajando en colaboración con el Hospital Francisco Lopez Lima de la Ciudad de General Roca, Río Negro.

La financiación de las líneas de investigación aquí presentadas provienen del proyecto de investigación de UNCOMA “04/F004: Informática Aplicada al Medio Académico y Científico” y el Inter-U “Colaboración UNComa-UNLP: docencia e investigación en Sistemas Paralelos” de PROMINF, y de otros proyectos de la UNLP y del grupo HPC4EAS.

2. Introducción

Nuestra investigación está enfocada en dos temáticas:

Consumo energético

Los sistemas de Cómputo de Altas Prestaciones (HPC, de *High Performance Computing*) han tenido, por décadas, el único objetivo de incrementar la velocidad de procesamiento de las aplicaciones científicas de alta demanda computacional. Esto se ve reflejado en la lista del TOP500 [2], que utiliza la métrica FLOPS (*F*loating-*p*oint *O*perations per *S*econd, cantidad de operaciones de coma flotante por segundo) para determinar el orden de clasificación de las supercomputadoras. Sólo importaban las prestaciones y, principalmente para el dueño de la supercomputadora, la relación precio/prestaciones. Así, las supercomputadoras se volvieron tan grandes que llegan a consumir tanta electricidad como una ciudad impactando fuertemente en lo económico (y posiblemente en lo ecológico y social, no solo por su consumo energético específico sino también por la energía requerida para refrigerar el sistema (que puede llegar a demandar tanto como el propio sistema de cómputo) y así asegurar su correcto funcionamiento.

El consumo energético, un factor limitante para el HPC [13], ha dado lugar a la nueva era de la supercomputación ecológica. Podríamos decir que el gran hito que determinó esta nueva era data de 2007. En ese año, una nueva entidad denominada Green500 [1], surgida como alternativa o complemento del TOP500, publica su primera lista que clasifica a las supercomputadoras de mayor eficiencia energética del mundo; esta clasificación considera el rendimiento por W (FLOPS/W) de cada supercomputadora al ejecutar un cierto benchmark. La eficiencia energética de los sistemas está mejorando drásticamente. A modo de ejemplo, basta observar que la primera supercomputadora publicada por el Green500 en la lista de Junio de 2014 procesaba 4389,82 GFLOPS/W y la de Noviembre de 2015 fue capaz de procesar 7031,58 GFLOPS/W. No obstante, como las mismas continúan incrementando su tamaño, su consumo energético sigue en aumento. Sin considerar el sistema de refrigeración, la actual supercomputadora más rápida del mundo demanda 17 MW de potencia, más que ninguna otra en la historia de la computación, significando un costo anual de aproximadamente 17 millones de dólares. Asimismo, la era exaescala aumentará notablemente la magnitud del problema energético.

Las diversas técnicas desarrolladas para mejorar la eficiencia energética de los procesadores involucran tanto al hardware como al software [7], y entre las más importantes se encuentran: (1) Explotación de paralelismo a través de múltiples cores que tienen bajas frecuencias de reloj, (2) Nivel de circuito y lógico [18] involucrando la tecnología de fabricación de los dispositivos (transistores) y reordenamiento de compuertas lógicas, (3) Interconexión de unidades funcionales y nodos involucrando rediseño de buses y redes de interconexión, (4) Optimizaciones de memoria en la jerarquía de memoria, (5) Arquitecturas adaptables e hibernación de recursos [9] involucrando cachés de activación parcial, discos, cores, unidades funcionales, tarjetas de interfaz de red, memorias, (6) Integración on-chip, (7) Procesadores de propósito especial involucrando el uso de Unidades de Procesamiento Gráfico (GPUs), (8) Rediseño de algoritmos de aplicaciones, (9) Planificación de tareas y asignación de tareas a recursos hardware [4], y (10) Escalado dinámico de frecuencia y tensión [17].

Aplicaciones para la salud

En un centro de salud hay diversas áreas en donde puede ser de gran utilidad los sistemas de cómputo de altas prestaciones. Algunas de estas áreas son las Unidades de Terapia Intensiva (UTI, o Unidad de Cuidados Intensivos - UCI) de adultos, pediátrica y neonatal. En una UTI, muchos sensores se conectan a los pacientes para medir distintos parámetros fisiológicos de alta frecuencia. Actualmente, se registra una medida de un conjunto de parámetros (no todos) de vez en cuando, posiblemente cada una o dos horas. Con la pérdida de estos datos, se están perdiendo muchas oportunidades para descubrir nuevos patrones de mediciones fisiológicas que podrían conducir a la detección temprana (predicción) de patologías. La detección temprana de patologías ofrece a los médicos la capacidad de planificar y comenzar tratamientos más pronto o potencialmente detener la progresión de una enfermedad, posiblemente, reduciendo la mortalidad y los costos. Los datos generados por los equipos médicos son un problema de grandes datos (*big data*) con restricciones de tiempo real para el procesamiento de algoritmos médicos diseñados para predecir patologías.

En la literatura existen algunos trabajos relacionados a la predicción de patologías como [14, 10, 12], sin embargo la mayoría son estudios médicos que no llegan a implementarse en sistemas de computación, a excepción del primero de ellos que se ha orientado a UTI neonatal.

3. Línea de investigación

Consumo energético

Nuestro trabajo no pretende desarrollar nuevas tecnologías hardware que tengan una mayor eficiencia energética (*diseño ecológico* [15]), sino gestionar mediante software el hardware existente para reducir el consumo energético (*uso ecológico* [15]). Nos centramos en el desarrollo de metodologías, modelos y construcción de software para administrar y gestionar el consumo de energía y prestaciones.

Nuestra investigación se enfoca en las siguientes temáticas energéticas aplicadas a sistemas de computación de altas prestaciones:

1. Predicción de energía y rendimiento. Es importante proveer a un administrador de sistema de herramientas que permitan predecir la energía y el rendimiento que producirían distintas configuraciones del sistema al ejecutar una dada aplicación paralela, y así poder seleccionar la configuración adecuada que mantenga el compromiso deseado entre tiempo de ejecución y eficiencia energética.
2. Gestión energética en mecanismos de tolerancia a fallos. La tolerancia a fallos agrega una carga de trabajo significativa al sistema de cómputo, sobre todo en sistemas que tienen enormes cantidades de unidades de procesamiento, haciendo necesario gestionar el consumo energético de los distintos mecanismos.

Aplicaciones para la salud

Esta línea está enfocada en el estudio y desarrollo de un sistema para detección temprana de patologías para la UTI, utilizando un sistema de cómputo de altas prestaciones.

Como un primer paso, hemos decidido comenzar a trabajar en el desarrollo de un sistema de monitoreo inteligente para la Unidad de Vigilancia Intermedia (UVI), del Hospital Francisco Lopez Lima de General Roca, Río Negro. La UVI de este hospital es una sala de internación (con 5 camas) donde el paciente está más controlado que en las salas de internación común y menos controlado que en las salas de terapia intensiva. En la UVI cada paciente tiene conectado un monitor médico, y hay enfermeros de forma permanente que controlan y asisten a los pacientes. Sin embargo, no hay médicos permanentes. Los médicos son llamados a demanda por los enfermeros cuando ellos detectan que su presencia es necesaria. El sistema de monitoreo inteligente que queremos desarrollar tiene como objetivo la disminución de muertes inesperadas. Debe

ser capaz de adquirir los datos fisiológicos de los monitores médicos y mediante la ejecución de algoritmos predefinidos detectar la necesidad de que algún paciente sea visitado por un médico o de notificar a los enfermeros de la realización de una cierta práctica sobre el paciente.

Contando ya con un subsistema de recuperación de datos de los equipos médicos que deberá desarrollarse en el primer paso, nos enfocaremos en el estudio de un sistema para detección temprana de patologías para la UTI.

4. Resultados y objetivos

Consumo energético

En referencia a la temática de caracterización energética de los sistemas, se ha estudiado la influencia de los modelos de programación paralela (OpenMP y MPI) y escalado de frecuencias de CPUs en sistemas de computación de altas prestaciones de memoria compartida [7]. También hemos estudiado los factores influyentes en el consumo energético de los sistemas de cómputo de altas prestaciones basados en CPUs y GPUs [8]. Posteriormente, hemos propuesto una metodología para caracterizar la potencia de sistemas de computación de altas prestaciones de memoria compartida [6]. La metodología involucra la búsqueda de factores influyentes en la potencia del sistema, realizando un análisis de sensibilidad de las propiedades de la carga de trabajo y parámetros del sistema en el comportamiento de potencia. La carga de trabajo considera aspectos de cómputo y comunicaciones de las aplicaciones. Esta metodología es similar a trabajos previos, pero nosotros proponemos un esquema de mayor profundidad que puede ayudar a mejorar la caracterización del sistema. También realizamos estudios sobre el impacto del sistema de Entrada/Salida en la eficiencia energética, que publicamos en [16].

En relación al diseño de algoritmos DVFS, hemos propuesto un algoritmo de mejora de la eficiencia energética para sistemas de memoria compartida [11], basado en la aceleración de los relojes de los cores en momentos de bajo paralelismo.

Respecto a la predicción de energía y estrategias energéticas en cloud computing, hemos propuesto una metodología para predecir el consumo energético de aplicaciones SPMD en entornos virtualizados, que publicamos en [4]. La metodología se basa en el desarrollo de un modelo analítico para predecir el consumo de energía y el producto energía-tiempo (EDP, Energy Delay Product). Las consideramos aplicaciones SPMD diseñadas con la librería de paso de mensajes

(MPI, Message Passing Interface), gran volumen de comunicación, que pueden generar problemas de balances de carga que afectan seriamente el tiempo de ejecución y la eficiencia energética. El método fue validado usando aplicaciones científicas y se observó una precisión de predicción entre el 91 y 96 %.

Los métodos de tolerancia a fallos tienen fuerte incidencia en el consumo energético de los sistemas de HPC, y resulta de suma importancia conocer, antes de ejecutar una cierta aplicación, el impacto que pueden producir los diferentes métodos y configuraciones del mismo. En [3], presentamos una metodología para predecir el consumo energético producido por el método de checkpoint coordinado remoto. La metodología se basa en una caracterización energética del sistema, una caracterización de la aplicación, y un modelo analítico que se instancia con los parámetros caracterizados. El modelo permite predecir la energía que consumirán los checkpoints para cualquier tamaño de problema y frecuencia de CPU de ejecución de checkpoints. Los resultados de las predicciones muestran una precisión mayor al 95 %.

Actualmente nuestros principales objetivos están orientados a:

- Proponer mecanismos de gestión de tolerancia a fallos que procuren un uso eficiente del cluster, permitiendo maximizar la productividad y minimizar el consumo energético.
- Mejorar la metodología propuesta de predicción del consumo energético de aplicaciones SPMD para permitir una mayor precisión de los resultados y extender la validación con otras aplicaciones.

Aplicaciones para la salud

Los resultados de esta línea de investigación actualmente se limitan a un análisis del estado general de las UTI del hospital Francisco Lopez Lima, que hemos publicado en [5]. Nuestros objetivos son:

- El desarrollo de un sistema de monitoreo inteligente para la UVI.
- El desarrollo de un sistema para detección temprana de patologías para la UTI.

5. Formación de recursos humanos

Los estudios aquí expuestos tienen como objetivo formar recursos humanos a nivel de grado y postgrado. Respecto al consumo energético,

en la Universidad Nacional del Comahue se ha desarrollado una tesis de grado, y se espera en 3 años más haber presentado una tesis doctoral relacionada a mecanismos de tolerancia a fallos que consideren el consumo energético. Respecto al tema de aplicaciones para la salud, se espera formar recursos humanos tanto a nivel de grado como postgrado.

Referencias

- [1] Sitio web del green500: <http://www.green500.org/>.
- [2] Sitio web del top500: <http://www.top500.org/>.
- [3] Javier Balladini, Marina Morán, Dolores Rexachs, and Emilio Luque. Metodología para predecir el consumo energético de checkpoints en sistemas de hpc. *XX Congreso Argentino de Ciencias de la Computación (CACIC 2014)*, 2014.
- [4] Javier Balladini, Ronal Muresano, Remo Suppi, Dolores Rexachs, and Emilio Luque. Methodology for predicting the energy consumption of spmd application on virtualized environments. *Computer Science and Technology (JCST)*, 13(3):130–136, 2013.
- [5] Javier Balladini, Claudia Rozas, Emmanuel Frati, Nestor Vicente, and Cristina Orlandi. Big data analytics in intensive care units: challenges and applicability in an argentinian hospital. *Computer Science and Technology (JCST)*, 2015.
- [6] Javier Balladini, Enzo Rucci, Armando Eduardo De Giusti, Marcelo Naiouf, Remo Suppi, Dolores Rexachs del Rosario, and Emilio Luque Fadón. Power characterisation of shared-memory hpc systems. In *XVIII Congreso Argentino de Ciencias de la Computación*, 2012.
- [7] Javier Balladini, Remo Suppi, Dolores Rexachs, and Emilio Luque. Impact of parallel programming models and cpus clock frequency on energy consumption of hpc systems. *Computer Systems and Applications, ACS/IEEE International Conference on*, 0:16–21, 2011.
- [8] Javier Balladini, Federico Uribe, Remo Suppi, Dolores Rexachs, and Emilio Luque. Factores influyentes en el consumo energético de los sistemas de computación de altas prestaciones basados en cpus y gpus. *XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011)*, pages 271–281, 2011.
- [9] Domingo Benitez, Juan C. Moure, Dolores Rexachs, and Emilio Luque. A reconfigurable cache memory with heterogeneous banks. In *Proceedings of the Conference on Design, Automation and Test in Europe, DATE '10*, pages 825–830, 3001 Leuven, Belgium, Belgium, 2010. European Design and Automation Association.

- [10] Andrea Bravi, Geoffrey Green, Andre Longtin, and Andrew J. E. Seely. Monitoring and identification of sepsis development through a composite measure of heart rate variability. *PLoS ONE*, 7(9), 2012.
- [11] Belén Casanova, Javier Balladini, Armando Eduardo De Giusti, Remo Suppi, Dolores Rexachs del Rosario, and Emilio Luque Fadón. Mejora de la eficiencia energética en sistemas de computación de altas prestaciones. In *XVIII Congreso Argentino de Ciencias de la Computación*, 2012.
- [12] Wei-Lung Chen and Cheng-Deng Kuo. Characteristics of heart rate variability can predict impending septic shock in emergency department patients with sepsis. *ACADEMIC EMERGENCY MEDICINE*, 2007.
- [13] Wu-Chun Feng, Xizhou Feng, and Rong Ge. Green supercomputing comes of age. *IT Professional*, 10(1), January 2008.
- [14] Carolyn McGregor. Big data in neonatal intensive care. *Computer*, 46(6):54-59, 2013.
- [15] San Murugesan. Harnessing green it: Principles and practices. *IT Professional*, 10:24-33, January 2008.
- [16] Javier Panadero, Sandra Méndez, Dolores Rexachs, Javier Balladini, Ismael Rodríguez, Adrián Pousa, Remo Suppi, and Emilio Luque. El impacto del sistema de entrada/salida en la eficiencia energética. In *XVII Congreso Argentino de Ciencias de la Computación*, 2011.
- [17] Barry Rountree, David K. Lownenthal, Bronis R. de Supinski, Martin Schulz, Vincent W. Freeh, and Tyler Bletsch. Adagio: making dvs practical for complex hpc applications. In *ICS '09: Proceedings of the 23rd international conference on Supercomputing*, pages 460-469, New York, NY, USA, 2009. ACM.
- [18] Vasanth Venkatachalam and Michael Franz. Power reduction techniques for microprocessor systems. *ACM Computing Surveys*, 237:2005, 2005.

Computación de Alto Desempeño y Datos Masivos: Arquitecturas, Modelos y Paradigmas.

Rubén Apolloni, Mercedes Barrionuevo, Mariela Lopresti, Natalia Miranda, Fabiana Píccoli, Marcela Printista, Cristian Tissera

LIDIC- Univ. Nacional de San Luís
San Luís, Argentina

{rubenga, mdbarrio, omlopres, ncmiran, mpiccoli, mprinti, ptissera}@unsl.edu.ar

Resumen

Este trabajo aborda las líneas de trabajo a seguir, con el objetivo de desarrollar e incluir técnicas de Computación de Alto Desempeño en problemas con datos masivos o problemas Big-Data. El desarrollo de la infraestructura, metodologías y herramientas de Computación de Alto Desempeño y su empleo eficiente en la solución de problemas que tratan grandes cantidades de datos, debe considerar la naturaleza de estos.

Palabras clave: Datos Masivos, Computación de Alto Desempeño, Arquitecturas Multicore y Many core

Contexto

Esta propuesta de trabajo se lleva a cabo dentro del proyecto de investigación “Tecnologías Avanzadas aplicadas al Procesamiento de Datos Masivos” y del proyecto binacional CAPG-BA 66/13 entre la Universidad Nacional de San Luis (UNSL) y la

Universidad Federal de Pernambuco (UFPE), Recife, Brasil.

El proyecto de investigación se desarrolla en el marco del Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC), de la Facultad de Ciencias Físico, Matemáticas y Naturales de la UNSL y el Centro de Informática de la UFPE.

Introducción

Cuando se habla de “big data” [MCJ13] se hace referencia a un conjunto de grandes volúmenes, diversos, complejos, longitudinales o distribuidos de datos, generados desde transacciones en Internet, sensores, instrumentos, videos, mails, redes sociales y desde una variedad de fuentes digitales disponible en la actualidad y futuras. Las técnicas tradicionales para el procesamiento, análisis y obtención de información útil

deben ser redefinidas para formular nuevas metodologías de abordaje. Trabajar con grandes volúmenes de datos implica grandes desafíos, debido a la necesidad de explorar un universo de nuevas tecnologías, las cuales no sólo hacen posible la obtención y procesamiento de los datos sino también realizan su gestión en un tiempo razonable [N13].

Para trabajar con datos a gran escala, es necesario tener en cuenta que la obtención de información útil será a partir de datos no estructurados como texto, audio, imagen y vídeo. Es por ello que se debe considerar la aplicación de nuevas metodologías para el procesamiento eficiente y eficaz de estos grandes volúmenes de datos. Además, como cada uno de los procesos involucrados en la obtención de la información a partir de dichos volúmenes implica un gran número de problemas computacionalmente costosos, el uso de nuevas técnicas y nuevas arquitecturas puede contribuir a mejorar su rendimiento; es por ello que la búsqueda y selección de técnicas de computación de altas prestaciones (HPC) en cada etapa o proceso involucrado permitirá resolver con eficiencia cada uno de los objetivos a plantearse.

Por todo lo expuesto, el procesamiento de grandes volúmenes de datos nos introduce en una nueva era de la computación, debido a que genera mayores demandas al procesador, a la memoria en todos los niveles (tanto a

memoria principal y memoria cache) [HP08], a los dispositivos de almacenamiento, y también requiere nuevas soluciones de software.

Entre las soluciones de software para este tipo de problemas se desarrollaron algoritmos y modelos de programación distribuidos y paralelos, como MapReduce[DG04], Hive[CWR12] e Impala[R13] que permiten procesar terabytes de información sin necesidad de cambiar las estructuras de datos subyacentes.

Entre los requerimientos de hardware, se encuentra la necesidad de mayor cantidad de almacenamiento para todos los datos, introduciendo nuevos desafíos de investigaciones y desarrollos. Además las aplicaciones requieren mayor capacidad de memoria, esperándose un aumento en la demanda. Otro aspecto a considerar es el consumo de energía, criterio muy importante a tener en cuenta en el diseño y desarrollo de sistemas de computadoras ya que está directamente relacionado con el consumo de energía total de la infraestructura computacional.

En la tecnología actual, la demanda de Computación de Alta Desempeño (HPC) se incrementa rápidamente. Con el continuo crecimiento de la Ley de Moore[G65], se observa una constante reducción del tamaño de los transistores, lo que permite empaquetar mayor cantidad de ellos en una única pastilla. Esto permite diseñar procesadores más potentes y/o incluir

mayor cantidad de núcleos dentro de una pastilla.

El continuo aumento en la cantidad de transistores posibilita el incremento del throughput, entre otras mejoras. Si bien este incremento permite construir procesadores más potentes. También, genera inconvenientes y limitaciones. Para continuar mejorando el desempeño del procesador, en la actualidad, se eligió diseñar procesadores de varios cores, más simples, dentro de una pastilla.

Con los sistemas de computación transformados desde procesadores de un solo núcleo a procesadores con numerosos núcleos, el paralelismo se hace omnipresente a todos los niveles. A nivel micro, el paralelismo es explotado desde los circuitos, el paralelismo a nivel de pipeline e instrucciones sobre procesadores multi-core. A nivel macro, se promueve el paralelismo desde múltiples máquinas en un rack a muchos rack en un centro de datos, hasta llegar a infraestructuras globales basadas en Internet [RR11].

La presente propuesta tiene como objetivo aplicar técnicas HPC en las etapas del proceso de obtención de información a partir de datos masivo considerando arquitecturas multi y many core como arquitecturas subyacentes.

Líneas de Investigación, Desarrollo e Innovación

Mejorar el trabajo con BigData implica considerar diferentes áreas, estas constituyen sendas líneas de investigación. Para lograrlo nos planteamos las siguientes líneas de investigación, ellas son:

- Modelos y paradigmas de computación de alto desempeño: la programación paralela involucra muchos aspectos, los cuales no se presentan en la programación convencional. El diseño de un sistema paralelo tiene que considerar entre otras cosas, el tipo de arquitectura sobre la cual se va a ejecutar el programa, las necesidades de tiempo y espacio requeridas por la aplicación; las técnicas y estructuras de programación paralela adecuadas para implementarla; y la forma de coordinar y comunicar las diferentes unidades computacionales dedicadas a resolver conjuntamente el problema. Además, en la última década las arquitecturas paralelas han evolucionado drásticamente (clusters de pc, procesadores multicore, procesadores manycore) existiendo un desafío adicional para las aplicaciones paralelas que consiste en explotar tales arquitecturas a su máximo potencial.
- Algoritmos y Estrategias de alta performance en grandes volúmenes de datos: un punto a tener en cuenta cuando lo que se desea es rendimiento, es el origen de los datos de un problema. Con el auge de Internet los datos son generados

automáticamente en forma distribuida (redes de sensores “wireless”, observaciones meteorológicas, centros de observaciones satelitales, etc). Otro inconveniente es el relacionado a la heterogeneidad de modos/formatos en la que se encuentra disponible la información y su administración eficiente.

- El análisis de las arquitecturas de procesadores y de las jerarquías de memoria juega un papel fundamental en determinar el desempeño global de un sistema HPC, y cobra mayor importancia dado que los volúmenes de datos cada vez requieren mayor capacidad de memoria y se espera que la demanda continúen aumentando. Una manera de abordar los problemas de densidad, consumo, desempeño y escalabilidad de las tecnologías de memorias y almacenamientos tradicionales, es empleando Memorias No Volátiles (NVM), en la actualidad se están desarrollando numerosas tecnologías NVM. A pesar de que se avizoran numerosos beneficios, también introducen nuevos desafíos, tales como, limitada durabilidad y alta latencia de las escrituras, que necesitan ser abordados. Otro de los aspectos que se abordará, es el manejo de la cache de último nivel (LCC), la cual es compartida por todo los núcleos del procesador, presenta el inconveniente que a medida que el número de núcleos aumenta, la contención causada por

las aplicaciones que comparten la LLC se incrementa, por lo que el rendimiento de estos sistemas estará muy influenciado por la eficiencia con la que se maneja la cache compartida.

En las líneas, las investigaciones tienen en cuenta la portabilidad de los desarrollos a pesar de las características propias de cada uno de los datos no estructurados.

Resultados y Objetivos

Como objetivos de las líneas de investigación nos planteamos facilitar el desarrollo de soluciones paralelas portables, de costo predecible y bajo consumo, capaces de explotar las ventajas de modernos ambientes de HPC a través de herramientas y “frameworks de computación” de alto nivel. Para ello será necesario proponer nuevas metodologías a ser aplicadas en cada una de las fases del tratamiento de datos masivos.

Formación de Recursos Humanos

Los resultados esperados respecto a la formación de recursos humanos son hasta el momento el desarrollo de 6 tesis doctorales y 4 tesis de maestría. Además se están ejecutando varias tesinas de grado.

Referencias

- [G65] G. E. Moore. Cramming More Components onto Integrated Circuits. Electronics 38, No. 8, pp 114-117 (April 19, 1965).
- [CWR12] [E. Capriolo](#), [D. Wampler](#), [J. Rutherglen](#). Programming Hive: Data Warehouse and Query Language for Hadoop. O'Reilly Media. 2012.
- [DG04] J. Dean and S. Ghemawat: MapReduce: Simplified Data Processing on Large Clusters. Proc. Sixth Symposium on Operating System Design and Implementation, 2004.
- [HP08] J. L. Hennesy and D. A. Patterson. Computer Organization & Design - The Hardware/Software Interface. Morgan Kaufmann, 4th edition, 2008.
- [HW11] G. Hager, G. Wellein. Introduction to High Performance Computing for Scientists and Engineers. Chapman & Hall/CRC Computational Science. 2011.
- [MCJ13] [V. Mayer-Schönberger](#), [K. Cukier](#). A.I. Jurado. Big data: La revolución de los datos masivos. Turner. 2013.
- [N13] J. Needham. [Disruptive Possibilities: How Big Data Changes Everything](#). Kindle Edition. O'Reilly Media Inc. 2013.
- [OPS01] S. Orlando, R. Perego, and F. Silvestri. Design of a Parallel and Distributed WEB Search Engine. In Proceedings of Parallel Computing (ParCo) 2001 conference. Imperial College Press, September 2001.
- [R13] J. Rusell: Cloudera Impala. O'Reilly Media, Inc. 2013.
- [RR11] T. Rauber, G. Runger. Parallel Programming for multicore and Cluster Systems. Springer. 2011.
- [W09] [T. White](#), [D. Cutting](#): Hadoop-The Definitive Guide. 2009 by O'Reilly Media 2009.

ESTUDIO Y DESARROLLO DE INTERFACES AVANZADAS ORIENTADAS A SISTEMAS DE ROBÓTICA

Walter A. Gemin, Raúl R. Rivera, Juana G. Fernández, Miguel A. Revuelta,
Melisa G. Kuzman, Roberto M. Hidalgo.

Laboratorio de Instrumentación Virtual y Robótica Aplicada (LIVRA) /
Departamento de Electrónica / Facultad de Ingeniería / Universidad Nacional de
Mar del Plata

J.B. Justo 4302, Tel.2234816600
gfernand@fi.mdp.edu.ar

Resumen

El término Robot [1] [2] se aplica a una variedad de sistemas que van desde los más simples manipuladores industriales, a los robots móviles "inteligentes", equipados con múltiples sensores y actuadores, capaces de adaptarse a diferentes entornos y moverse en forma autónoma en ellos. Los desarrollos en este campo abarcan una amplia gama de especialidades, desde la mecánica a la electrónica e informática. Dentro de la especialidad electrónica y en común con la Instrumentación Virtual, involucra el conocimiento y aplicación de sensores, actuadores, software de procesamiento e interfaces gráficas para la supervisión y comando remoto.

Este Proyecto tiene como objetivo estudiar y desarrollar nuevas tecnologías en percepción en tiempo real, seguimiento y comando desde dispositivos móviles, interfaces basadas en BCI, navegación y control de tracción inteligente.

Como resultado, se esperan obtener interfaces y algoritmos de procesamiento

para plataformas portátiles basadas en Android y módulos Arduino

Palabras clave: Robótica, Cinemática, Lenguaje C, Android, Arduino.

Contexto

El Proyecto en el que está inserta la línea de investigación se denomina: Instrumentación virtual, estudio y desarrollo de interfaces avanzadas orientadas a sistemas de robótica.

La institución que acredita el Proyecto es la Secretaría de Ciencia y Tecnología de la UNMdP (Universidad Nacional de Mar del Plata).

El Proyecto está financiado principalmente por la UNMdP, con aportes de la Organización Techint mediante la iniciativa "Actualización en Docencia e Investigación en Instrumentación Virtual y Robótica", el Ministerio de Educación de la Nación a través del programa de articulación con las escuelas de enseñanza media: "Jugando con robots en el aula" y la SPU del Ministerio de Educación. Mediante el

programa "Universidad, Diseño y Desarrollo Productivo "con el título: "UMDP-4 / Interface para silla de ruedas" [4].

Introducción

La Instrumentación Virtual y la Robótica son las áreas con mayor crecimiento en la ingeniería. Los desarrolladores son, entre otros, Ingenieros Electrónicos y en Computación, cuya actividad es el diseño de prototipos y equipos, integrando tecnologías de sensores y actuadores, algoritmos de cálculo y comunicación, sistemas autónomos inteligentes y el hardware integrado. Son tecnologías multidisciplinarias que hacen uso de recursos de otras ciencias afines, permitiendo obtener sistemas flexibles que se adaptan a las diferentes tareas de producción.

La Instrumentación Virtual esta enfocada al desarrollo de los instrumentos encargados de medir señales, registrar datos y efectuar las acciones de control, para lo cual se requiere de una etapa de actuación, que conforma la interface entre la computadora y el sistema a controlar, mediante drivers de potencia. El usuario final del sistema de instrumentación interactúa con la representación gráfica de los indicadores y botones de control virtuales en la pantalla de una PC.

Respecto de la Robótica, existen diferentes tipos robots y el mas usado en la industria es el manipulador [7] que se puede describir como sistema automático servo-controlado, reprogramable, polivalente, capaz de posicionar y orientar piezas, útiles o dispositivos especiales, siguiendo trayectorias variables reprogramables, para la ejecución de tareas variadas. Su unidad de control incluye un dispositivo de procesamiento y

sensores de percepción del entorno. Normalmente realiza una tarea de manera cíclica, pudiéndose adaptar a otra sin cambios permanentes en su material.

De esta forma cada robot se transforma en una célula flexible que disminuye el ciclo de trabajo de un producto, puede operar en forma autónoma o interconectarse con otros para automatizar procesos complejos.

Este Proyecto tiene como objetivo estudiar y desarrollar nuevas tecnologías aplicadas a instrumentación virtual y robótica, en las siguientes temáticas:

Brazos robóticos: de gran utilidad en la industria y cuya constitución guarda similitud con la anatomía de las extremidades superiores del cuerpo humano. Para hacer referencia a los distintos elementos que lo componen se usan términos como codo, muñeca, mano etc. El brazo está formado por una serie de elementos estructurales sólidos unidos mediante articulaciones que permiten un movimiento relativo entre dos eslabones consecutivos. Cada articulación provee al robot de un grado de libertad lo que le permite realizar movimientos horizontales o verticales. Estos movimientos están generados por servomotores ubicados en cada una de las articulaciones. El trabajo del brazo robot normalmente se limita a pocos movimientos repetitivos de sus ejes. Éstos son casi siempre 3 para el brazo propiamente dicho y 3 para la mano o puño y su radio de acción queda determinado por un espacio tridimensional que cubre una semiesfera.

Conociendo la cinemática y la dinámica de un manipulador con elementos serie, es de interés saber cómo controlar los actuadores de sus articulaciones para cumplir una tarea deseada según una determinada trayectoria, libre de posibles

obstáculos. El problema de control de un manipulador se puede dividir en el problema de planificación y el control del movimiento. El movimiento del brazo de un robot se suele realizar en dos etapas. La primera de ellas es la del control del movimiento de aproximación del brazo a lo largo de una trayectoria planificada, desde una posición inicial hasta la vecindad de la posición de destino. La segunda es la de control del movimiento fino en el cual el efector final del brazo interacciona dinámicamente con el objeto para completar la tarea, utilizando información obtenida a través de la realimentación sensorial.

La utilización de sensores externos permite a un robot interactuar con su entorno de una manera flexible, en contraste con las operaciones preprogramadas en las cuales a un robot se le enseña para efectuar tareas repetitivas mediante un conjunto de funciones almacenadas en una memoria. Aunque esto último es la forma predominante de operación de los robots industriales actuales, la utilización de tecnología sensorial para dotar a las máquinas con un mayor grado de “inteligencia”, al tratar con su entorno, es tema de investigación y desarrollo activo en el campo de la robótica, dentro del cual trabajar en este Proyecto.

Sistemas autónomos: el continuo desarrollo de vehículos autónomos [8] [9] ha otorgado una mayor importancia a los sistemas de visión por computadora [10] [11] [12]. Muchos modelos de automóviles ya incluyen sensores y cámaras para la asistencia al conductor en el estacionamiento, prevención de accidentes y choques. En los sistemas robóticos autónomos son imprescindibles los métodos de detección y seguimiento de objetos, ya que estos se utilizan para

evitar accidentes, reconocer objetivos o efectuar tareas de monitoreo dinámicas. Paralelamente, ha surgido en los últimos años un creciente interés por la introducción de la robótica en la educación. Este fenómeno se debe en parte a la aparición de distintas plataformas robóticas de bajo costo que permiten acercarla a un tipo de público que antes no tenía acceso a ella. Ejemplo de estas plataformas puede ser Arduino, una placa electrónica de código abierto que permite implementar de forma sencilla el control de motores y sensores.

Este Proyecto une la robótica y la visión por computadora para crear sistemas autónomos más efectivos que puedan operar en determinados entornos realizando tareas predeterminadas.

Se plantea diseñar un sistema que una la flexibilidad de una plataforma robótica con la potencia de un dispositivo móvil. Esto implica trabajar con tecnologías diferentes en un mismo sistema: dispositivos móviles con Android y Arduino. Por lo tanto, el robot realiza la detección y el seguimiento de un objetivo mediante algoritmos de visión y debe ser capaz de alternar distintos comportamientos para desplazarse hacia su objetivo y evitar los obstáculos cuando sea necesario.

Tracción Inteligente: estudiar las técnicas de tracción y dirección inteligentes de un robot móvil con ruedas y desarrollar los algoritmos y generar las señales de control para los motores que permitan optimizar el comportamiento de la configuración propuesta.

En general los robots móviles distribuyen sus sistemas de tracción y dirección sobre los ejes de sus ruedas de acuerdo a las exigencias de velocidad, maniobrabilidad y características del terreno. La precisión

y rapidez con que el robot móvil debe alcanzar su destino, implica tener un sistema de tracción y dirección confiables, que asegure su maniobrabilidad. Esta confiabilidad y maniobrabilidad, determinan las características del sistema de tracción y dirección, no sólo en lo que respecta a la técnica, sino también al número de ruedas necesarias y al tipo y disposición de éstas para lograr una estructura mecánica estable. Existen tres sistemas básicos a partir de los cuales se pueden obtener este comportamiento: Tracción y dirección en ejes independientes, en un mismo eje y sobre todos los ejes.

El sistema de tracción y dirección no sólo está relacionado con la disposición de ruedas adoptada, sino que también depende de los algoritmos de control local de los motores y la mecánica asociada a estos. A medida que se requiere, más confiabilidad (alta maniobrabilidad, máxima tracción en sus ruedas motorizadas y adherencia de todas sus ruedas, etc.) la mecánica, electrónica e informática asociadas son más complejas.

Es importante puntualizar que el control de la tracción es un aspecto crítico en los robots móviles que se desplazan sobre superficies con un bajo coeficiente de fricción estática, o que deben moverse sobre suelos poco compactos (arena, lodo, nieve), donde un deslizamiento excesivo puede provocar incluso que las se hundan en el terreno y queden atrapadas en el mismo. Lo expuesto hasta aquí justifica la importancia del estudio y ensayos de las características del sistema de tracción en un robot móvil.

Líneas de Investigación, Desarrollo e Innovación

Las líneas de investigación abarcadas son:
Estudio de la cinemática directa e inversa de un robot, basado en el algoritmo de Denavit- Hartenberg.

Estudio del microcontrolador PIC y módulos Arduino, mediante los entornos de desarrollo en lenguaje C, PIC CCS y Processing.

Desarrollo y simulación en Matlab, de algoritmos de cálculo de movimiento y posicionamiento.

Adaptación de los algoritmos a los microcontroladores PIC y módulos Arduino.

Desarrollo de software e interfaces de comunicación serie de los microcontroladores PIC y módulo Arduino.

Desarrollo de aplicaciones para PC y dispositivos portátiles como notebook, tablets o celulares, para el comando y pruebas de funcionamiento de sistemas robot.

Estudio de las técnicas que se emplean en la visión por computadora: procesamiento de imágenes, detección de características, reconocimiento de objetos y análisis de movimientos.

Desarrollo librerías para detección y seguimiento de objetos desde una señal de video.

Pruebas de las tecnologías desarrolladas en módulos de robots experimentales.

Resultados y Objetivos

Resultados:

Como resultados de recientes trabajos de extensión, relacionados con los temas del Proyecto pueden mencionarse:

Desarrollo y optimización de algoritmos de cálculo de movimientos, basados en microcontroladores de bajo costo, con interfase de comunicación y salida PWM.

Desarrollo de sistemas de potencia con entrada PWM y salida a motores de continua.

Pruebas y optimización del sistema orientado al uso de material de bajo impacto ambiental y elevada autonomía [3].

Desarrollo de interfaces para control por Joysticks potenciométricos, inalámbricos y virtuales [4].

Desarrollo de sistema de control de movimiento basado en Arduino y comandado mediante joystick virtual en Android, para el control remoto inalámbrico, por medio de Bluetooth, de una silla de ruedas eléctrica [5] [6].

La producción en investigación relacionada con el Proyecto se resume en: 6 publicaciones internacionales, 5 presentaciones a congresos nacionales y 10 internacionales, 3 premios y 1 patente.

Los integrantes del Proyecto tienen a su cargo el dictado de asignaturas muy relacionadas con la temática propuesta:

- Tratamiento Digital de Señales.
- Sistemas Digitales
- Instrumentación Electrónica.
- Instrumentación Virtual.
- Robótica Aplicada.
- Laboratorio I – Ing. en Materiales
- Programación Estructurada
- Programación Orientada a Objetos

Además debe mencionarse el dictado de diversos cursos de extensión en temáticas que abarcan: Sensores, Adquisición Digital de Señales, Instrumental Electrónico, Microcontroladores, Lenguaje de Programación C y Compiladores Cruzados, Capacitación en redes LAN y WAN, Programación en Matlab y en Android.

Objetivos:

Fomentar la generación de conocimientos y recursos humanos dentro de un ámbito de investigación, para difundir y promover el desarrollo y producción nacional de sistemas de Instrumentación Virtual y Robótica.

Formación de Recursos Humanos

El equipo de trabajo está constituido por 7 investigadores categorizados en el programa de incentivos y 3 investigadores en formación. Todos ellos se desempeñan como docentes del Dpto. de Ingeniería Electrónica y Computación y desarrollan sus actividades de investigación en el LIVRA (Laboratorio de Instrumentación Virtual y Robótica Aplicada).

Becarios Alumnos:

Los alumnos pertenecientes a la carrera Ingeniería Electrónica, que participan en este Proyecto, se distinguen en las categorías de avanzados: 3 alumnos de quinto año y de iniciación: 3 alumnos de tercer año, dentro del programa "Universidad, Diseño y Desarrollo Productivo" con el título: "*UMDP-4 / Interface para silla de ruedas*", aprobado y otorgado, por la SPU del Ministerio de Educación.

Otros 4 alumnos avanzados desarrollaron los siguientes temas:

Desarrollo de un software de configuración para redes inalámbricas bajo protocolo ZigBee orientadas a aplicaciones industriales.

Sistema robótico de control y monitoreo remoto, de aplicación en la industria del Oil&gas.

Sistema de visión robótica para dispositivos móviles implementado en Android.

Estudio y experimentación de las técnicas de tracción y dirección inteligentes de robots móviles.

Tesis de grado

Durante los últimos tres años, en el marco de los temas involucrados en este Proyecto, se han ejecutado 20 Tesis de Grado pertenecientes a la carrera Ingeniería Electrónica. La mayoría de estas tesis se realizaron en respuesta a demandas concretas del medio. Un listado completo puede obtenerse a través de la página web del Departamento de Ingeniería Electrónica [13].

Cursos de postgrado

Dictado del curso de postgrado: “Procesamiento Digital de Señales” que actualmente forma parte del plan de cursos del Doctorado en Ingeniería mención Electrónica y se dicta en forma regular cada 2 años desde 1997.

Cursos de especialización

Introducción y Aplicación del Software de Procesamiento Matlab
Tipo de actividad: Curso de Especialización, Personal del INIDEP.

Referencias

1. CRAIG, John J. Robótica, Pearson Education 3era Ed. 2006.
2. BARRIENTOS, Antonio. Fundamentos de Robótica, McGraw-Hill/ Interamericana, 2007.
3. GEMIN W. A., RIVERA R. R., FERNÁNDEZ J. G., HIDALGO R. M., REVUELTA M. A., URIZ A. J. “Desarrollo de un sistema de control y potencia de bajo costo para sillas de ruedas motorizadas”. JAIIO 42 – AST 2013.
4. <http://diseñoydesarrollo.siu.edu.ar/int erfaseparasilladeruedas.html>
5. GEMIN W. A., RIVERA R. R., FERNÁNDEZ J. G., HIDALGO R. M., REVUELTA M. A., URIZ A. J. “Control de Sillas de Ruedas Eléctricas Mediante Joystick y Celular”. INNOVAR 2013.
6. <http://www.youtube.com/watch?v=ltlj7qz9ksw>
7. KOIVO A. J. Fundamentals for Control of Robotic Manipulators. Wiley, 1989.
8. ASTUA C., BARBER R., CRESPO J. y JARDON J. “Object Detection Techniques Applied on Mobile Robot Semantic Navigation”. SENSORS. www.mdpi.com/journal/sensors.2014.
9. BUCHLI J. Mobile robotics, moving intelligence. Ed. VERLAG 2006.
10. SOLARI F., CHESSA M., SABATINI S. Machine Vision - Applications and Systems. Ed. INTECH 2012.
11. KYPRAIOS I. Advances in Object Recognition Systems. Ed. INTECH 2012.
12. OBINATA G., DUTTA A. Vision systems applications. Ed. I-TECH 2007.
13. www3.fi.mdp.edu.ar/electronica/Tesis_Grado.htm

Análisis de Señales Acústicas para dispositivos de personas con discapacidad, aplicando Teoría de Procesamiento de Radar

Sáenz Fabián¹, Romero Carlos²,
Bernal Paúl

Departamento de Eléctrica y
Electrónica / Universidad de las
Fuerzas Armadas ESPE
Sangolqui, Av. Progreso s/n, Ecuador,
+5933989400 ext. 1865
fgsaenz@espe.edu.ec /
cgromero@espe.edu.ec /
cpbernal@espe.edu.ec

Resumen

El presente artículo establece las bases para la implementación de un arreglo de micrófonos empleando la teoría LCMV y la técnica de Beamforming, para evaluar el mejor algoritmo que existe dentro de las diferentes familias, empleando como medidas de desempeño el error cuadrático medio y las propiedades de los algoritmos LCMV, además de configurar en diferentes estructuras para verificar la versatilidad de los mismo dentro de diferentes ambientes de trabajo u aplicaciones. Con la idea de una implementación de bajo costo para personas con capacidades especiales en cuanto a la pérdida del sentido de audición y mejorar su vida para incluirlos en la sociedad en general.

Palabras clave: Recursive Least Square, Beamforming, Linear

Prediction Coefficients, Señales Sonoras.

Contexto

El proyecto se encuentra enmarcado en la línea de investigación de Procesamiento de señales y Sistemas de Tiempo Real, con especial énfasis en las ayudas tecnológicas para discapacitados. Se trabaja en la Universidad de las Fuerzas Armadas ESPE, con la colaboración de la Universidad Nacional de La Plata, donde nos encontramos realizando el Doctorado en Ciencias Informáticas.

I. INTRODUCCIÓN

Existe un sin número de técnicas de procesamiento de señales sonoras dependiendo de la aplicación que se pretenda realizar, en este caso la aplicación en un arreglo de micrófonos que me permite: mejorar la señal, realzar la señal, calcular la dirección de la fuente sonora, bloquear señales no deseadas entre otros.

Para el presente proyecto lo que se buscara es sentar las bases del análisis enfocado en técnicas de Beamforming substituyendo las antenas por arreglos de micrófonos de forma lineal y las ondas electromagnéticas por señales sonoras. Existen algunos algoritmos para realizar el procesamiento MUSIC, CAPON, LMMSE, etc. Para este estudio emplearemos los últimos LMMSE "Linear Minimum Mean Square Error" por concentrar los algoritmos a emplearse en nuestro estudio, del cual extraeremos el receptor óptimo. Del estudio de los algoritmos del tipo LMMSE desglosaremos los del tipo filtros adaptativos usando la técnica conocida como LCMV "Linearly-Constrained Minimum Variance".

II. CONCEPTOS BÁSICOS

En esta sección abordaremos los

¹ Estudiante de Doctorado en Ciencias de la Universidad Nacional de La Plata UNLP

² Estudiante de Doctorado en Ciencias de la Universidad Nacional de La Plata UNLP

criterios básicos tanto para la geometría del arreglo, como a la familia de algoritmos a ser seleccionadas. Además de eso una explicación de cómo se procesa la señal sonora, esto detallamos da siguiente forma:

A. Geometría del Arreglo

La geometría del arreglo tiene que ver con la figura geométrica con la que se van a disponer los sensores que para este caso son micrófonos pudiendo ser: lineales, circulares bidimensionales, tridimensionales, etc. Para este estudio se implementó un arreglo lineal uniforme como se muestra en la Figura 1.

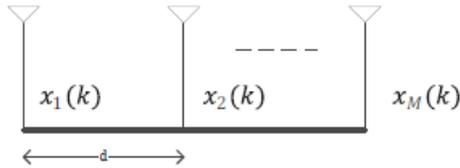


Fig. 1. Arreglo Lineal Uniforme.

B. Beamforming Adaptativo

El *Beamforming* Adaptativo surge cuando no se conoce las propiedades estadísticas de las señales, y se emplea para el diseño las estadísticas recogidas en línea, además de permitir varias familias de algoritmos en su estructura haciendo le más versátil su implementación. La estructura típica de esta estructura se observa en la Figura 2.

Se pretende establecer que:

$$\min_w \xi(k) \text{ sujeto a } C^H w = f \quad (1)$$

Dónde: $\xi(k) = E[|e(k)|^2]$

Con la matriz de restricción C de dimensión $MN \times p$ y el vector de ganancia f de dimensión $p \times 1$, siendo que el óptimo Beamforming es:

$$w(k) = R^{-1} p + R^{-1} C (C^H R^{-1} C)^{-1} (f - C^H R^{-1} p) \quad (2)$$

Donde $R = E[x(k)x^H(k)]$ y $p = E[d^*(k)x(k)]$.

Con

$$w(k) = [w_1^T(k) \ w_2^T(k) \ \dots \ w_M^T(k)]^T, \quad x(k) = [x_1^T(k) \ x_2^T(k) \ \dots \ x_M^T(k)]^T \text{ y } x_1^T(k) = [x_1(k) \ x_1(k-1) \ \dots \ x_1(k-N+1)]$$

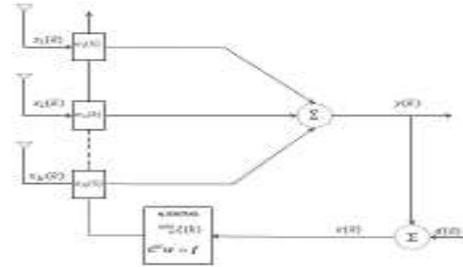


Fig. 2. Arreglo Beamforming

En ausencia de información estadista, se puede escoger que:

$$\min_w \left[\xi(k) = \sum_{i=0}^k \lambda^{k-i} |d(i) - w^H x(i)|^2 \right] \quad (3)$$

Sujeto a $C^H w = f$ con $\lambda \in (0, 1]$, lo que se aplica a la solución

$$w(k) = R^{-1}(k)p(k) + R^{-1}(k)C (C^H R^{-1}(k)C)^{-1} (f - C^H R^{-1}(k)p(k)) \quad (4)$$

Donde $R(k) = \sum_{i=0}^k \lambda^{k-i} x(i)x^H(i)$

y $p(k) = \sum_{i=0}^k \lambda^{k-i} d^*(i)x(i)$

La función de costo como alternativa más adecuada es

$$\min_w [\xi(k) = \|w(k) - w(k-1)\|^2 + \mu |e(k)|^2] \quad (5)$$

Sujeto a $C^H w = f$, lo que da como solución

$$w(k) = P[w(k-1) + \mu e^*(k)x(k)] + F \quad (6)$$

donde $e(k) = d(k) - w^H(k-1)x(k)$, μ una constante positiva pequeña llamada paso de adaptación, $P = C(C^H C)^{-1}C^H$ y $F = C(C^H C)^{-1}f$.

Para solventar problemas de velocidad de convergencia se puede establecer que

$$\min_w [\xi(k) = \|w(k) - w(k-1)\|^2] \quad (7)$$

Esto en relación con:

$$X^T(k)w^*(k) = d(k) \\ C^H w(k) = f$$

Dando como resultado;

$$w(k) = P[w(k-1) + \mu X(k)t(k)] + F \quad (8)$$

donde $e(k) = d(k) - XH(k)w^*(k - 1)$
 y $t(k) = XH(k)PX(k) - 1 e^*(k)$

C. Familia de Algoritmos

Los algoritmos del tipo LCMV podemos dividir en tres grupos, los con restricción como el caso CLMS, que se estableció sus parámetros en la subsección anterior la estructura GSC que permite que cualquier algoritmo unconstrained pueda ser empleado como los algoritmos con restricción, como se observa en la Figura 3. Con el objetivo de resolver $R = wb$, con la premura que dependerá de la matriz R, matriz de correlación de los datos del vector de entrada $x(k)$, y b el vector de correlación cruzada entre la señal de entrada y la respuesta deseada $d(k)$ y la estructura Householder que mediante la, matriz Q y los reflectores de Householder cumple como el GSC pero de una manera más eficiente computacionalmente hablando.

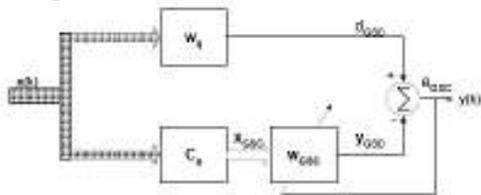


Fig. 3. Estructura GSC.

Al Householder tenemos que hacer una analogía con la estructura GSC esto lo podemos observar en la Figura 4. donde comparamos el vector w con la matriz Q de la transformación.

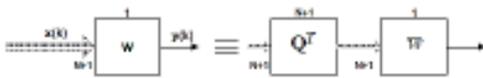


Fig. 4. Comparación con la transformación Householder.

Una vez que se ha conseguido dicha partición podemos asemejarlo a la estructura GSC con las mismas consideraciones que se emplearon para redefinir las ecuaciones del modelo GSC, esta forma de asemejarlo lo podemos ver en la Figura 5.

Como se observa en la Figura 3 es muy similar a la estructura de la Figura 5, para determinar cuál de estas estructuras resulta más eficiente las evaluaremos de acuerdo a los siguientes parámetros: Velocidad de convergencia, Estabilidad, Misajustment y Costo computacional

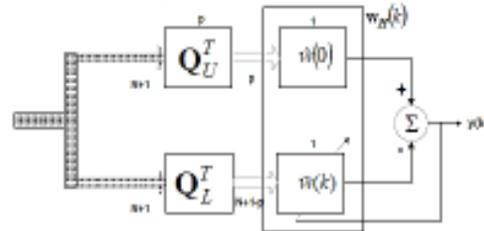


Fig. 5. Desglose de las matrices estructura Householder.

III. ANÁLISIS DE RESULTADOS

Para verificar el desempeño de los algoritmos versus los parámetros indicados en la Sección II, para eso utilizaremos un arreglo lineal uniforme con $M=4$, esto quiere decir que el número total de micrófonos empleados es 4, con el fin de disminuir el costo de implementación a futuro de un prototipo.

Para que los resultados obtenidos sean adecuados analizaremos primero los micrófonos para determinar si su respuesta es igual o al menos parecida puesto que en la teoría los sensores tienen las mismas características, se empleara una señal de ruido blanco como señal patrón ya que la teoría estadística, dicho ruido se encuentra presente en todo el espectro de frecuencia y es estacionario en sentido amplio, los resultados se pueden observar en la Figura 6.

La respuesta no es igual, sobre todo en las altas frecuencias, pero como nuestro análisis es las bajas frecuencias 4kHz señal voz no se tiene alto impacto. Para mejorar el desempeño, se emplea la estructura como un identificador de planta o

sistema y modelamos dichas distorsiones poniendo como referencia al Mic 1, dichas mejoras se las puede observar en la Figura 7:

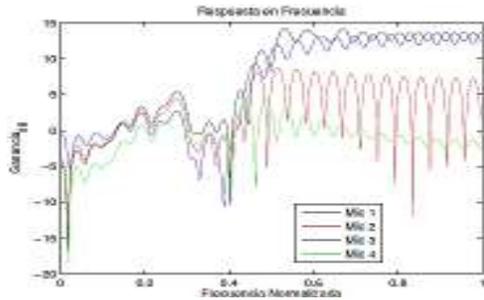


Fig. 6. Respuesta en Frecuencia.

Como se observa hay una mejora y en este caso podemos decir que la respuesta de los micrófonos se aproxima entre ellos, de esta manera

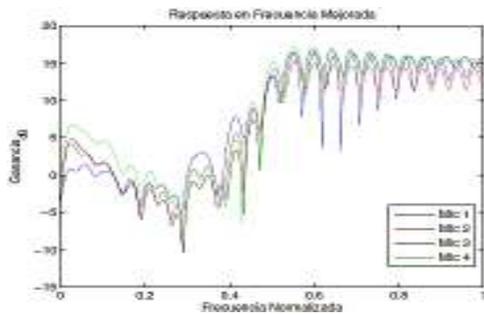


Fig. 7. Respuesta en Frecuencia Mejorada.

podemos mejorar los resultados en las aplicaciones que se pretenda emplear. Otra característica que se presentan en las señales es el desfase fruto del hardware y la misma propagación de la señal sonora que dependiendo de la aplicación puede ser crítica o no esto lo podemos observar en la Figura 8

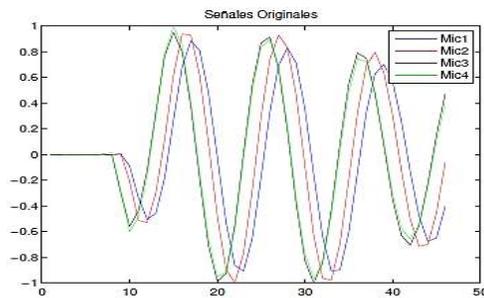


Fig. 8. Señales Originales.

De igual manera se emplea los algoritmos como predictores y podemos obtener una mejora en la señal que permita compensar dicho desfase esto lo observamos en Figura 9. Nuevamente se observa una mejoría en la señal que a futuro traerá mejores resultados en cuanto a simulaciones y demás señales con las fases corregidas

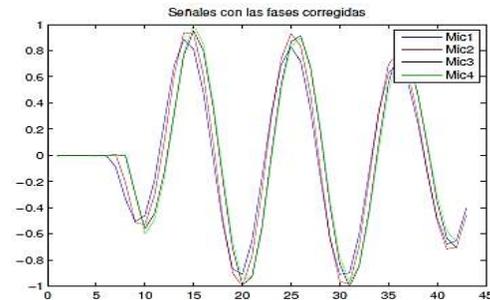


Fig. 9. Señales Corregidas

No se debe de olvidar que es una onda propagando se en el medio la cual también puede ser ecualizada si fuera necesario dependiendo como se recalcó de la aplicación esta ecualización puede permitimos mejoras en cuanto a comparación de señales, dicho proceso lo podemos observar en la Figura 10.

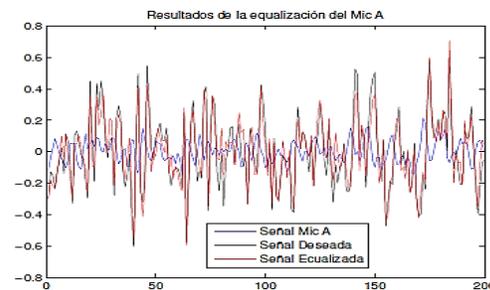


Fig. 10. Señal Ecualizada.

Se puede observar la ecualización del micrófono 1, ahora se tendría que definir cuál de todas las familias de algoritmos es el más idóneo para la implementación en miras de hacerlo en un FPGA donde lo que primaria es un bajo costo computacional para de esta manera ahorrar en la construcción de un

prototipo. Podemos definir qué familia resulta más idónea a través de una curva del error cuadrático medio o MSE, donde podríamos determinar la velocidad de convergencia de las diferentes familias como se muestra en la Figura 11

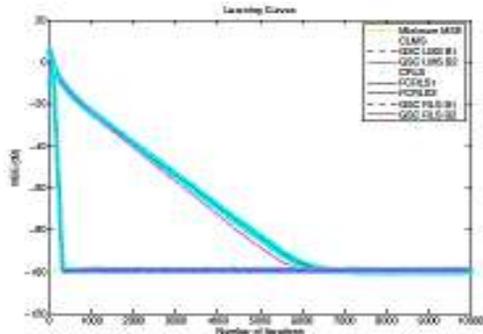


Fig. 11. MSE.

Todos los algoritmos convergen para el error mínimo que es de -100dB impuesto para verificar su desempeño lo cual establece que el desajuste (misajustement) entre el error teórico y el simulado u obtenido, varia muy poco una vez que hayan convergido, pero si se puede establecer que los que tienen una estructura recursiva convergen más rápida como lo son los RLS, pero a cambio tienen un costo computacional más elevado puesto que en sus estructuras contemplan inversión de matrices.

IV ANÁLISIS DE RESULTADOS

De las diferentes figuras mostradas podemos establecer que los algoritmos LCMV son muy versátiles en diferentes aplicaciones ya sea identificación, predicción, ecualización, etc. Todos convergen con diferentes velocidades, más rápidos, más complejos computacionalmente hablando, pero todo esto enfocado en una implementación práctica en FPGA. Una forma de optimizar dicha implementación podría ser emplear estructuras de filtros transversales y de esta manera reducir el número de multiplicaciones, sumas restas y

divisiones y omitir la inversión de matrices.

BIBLIOGRAFÍA

- [1] Mallat S. G.; Zhang Z.; Matching Pursuits With Time-Frequency Dictionaries, IEEE Transactions on Signal Processing, pp 3397-3415, Vol 41, No 12 December 1993.
- [2] Schetzen M.; The Volterra and Wiener Theories of Nonlinear Systems, Jhon Wiley & Sons, Inc., 1980.
- [3] Haykin S.; Redes Neurais: Principios e Prática, 2. Ed., Porto Alegre, RS, Bookman, 2001.
- [4] Quatieri, T.F.; Reynolds D.A.; O'Leary G.C; Estimation of Handset Non-linearity with Application to Speaker Recongition, IEEE Transactions on Speech and Audio Processing, pp 567-584, Vol 8, Nr 5 September 2000.
- [5] Reynolds D.A, HTIMIT and LLHDB: Speech Corpora for the Study of Handset transducer Effects, ICASSP, 1997.
- [6] Dimolitsas, S., Objective speech distortion measures and their relevance to speech quality assessments, IEE proceedings, pp317-324, Vol 136, Pt 1, Nr 5, October 1989.
- [7] Reynolds, D.A, Zissman M.A, Quatieri T.F., O'Leary, G.C., Carlson B.A., The effects of telephone transmission degradations on speaker recognition performance, ICASSP, 1995.
- [8] V. John Mathews. Adaptive Polynomial Filters, IEEE Signal Processing Magazine, vol 6, pp. 10-26, 1991.
- [9] Markel J. D., Gray A. H. Jr. Linear Prediction of Speech Springer-Verlag Berlin Heidelberg, 1976.
- [10] Burrus C. S.; Gopinath R. A. Gua H. Introduction to Wavelets and Wavelet Transforms - A Primer, Prentice Hall New Jersey, 1998.
- [11] Daubechies, I. Ten Lectures

on Wavelets, SIAM, Philadelphia.
PA, 1992. Notes From the 1990 CBMS-
NSF Conference on Wavelets and
Applications at Lowell, MA.

Control programático de Drone DJI Phantom 3

Gonzalo Zabala, Matías Teragni, Sebastián Blanco, Ricardo Morán

CAETI /Facultad de tecnología informática / Universidad Abierta Interamericana

Montes de Oca 745, Capital Federal, Argentina, (5411)4301-5240

Gonzalo.zabala@uai.edu.ar, Matias.teragni@uai.edu.ar, sebastian.blanco@uai.edu.ar,
Ricardo.moran@uai.edu.ar

Resumen

Este artículo trata sobre las limitaciones y los desarrollos realizados para poder tomar control de un Drone Phantom 3 de la empresa DJI, con el objetivo de desarrollar futuro comportamiento autónomo en el mismo.

Palabras clave: Phantom3, Drone, Control desde PC, Vuelo Autónomo

Contexto

El presente proyecto será radicado en el Centro de Altos Estudios en Tecnología Informática (CAETI), dependiente de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana. El mismo se encuentra inserto en la línea de investigación “Sociedad del conocimiento y Tecnologías aplicadas a la Educación”. El financiamiento está dado por la misma Universidad Abierta Interamericana

Introducción

La empresa de tecnología china DJI, actualmente la líder del mercado en drones comerciales y recreacionales, alcanzó esta posición principalmente gracias al Phantom 2.

Este dispositivo es un drone relativamente económico, sumamente simple de utilizar, ideal para aficionados. Cuenta con una buena autonomía considerando su tamaño y costo. La existencia de este tipo de robots comerciales ahorra los costos y complejidades inherentes al desarrollo individual y permite a las personas realizar grabaciones y fotografía aérea sin la necesidad de tener las habilidades y conocimientos técnicos necesarios para generar sus propios drones. En particular el Phantom 2 fue el drone más vendido a nivel mundial hasta la fecha. (Kakaes, 2015)

Esta empresa lanzó un nuevo modelo, llamado Phantom 3, un drone económico capaz de ser controlado a una distancia de 3.5km, con una autonomía de más de 20 minutos, y equipado con una cámara estabilizada de 4k que permite sacar tanto fotos como video en alta definición. A su vez, cuenta con la capacidad de ser controlado manualmente, o de movilizarse de forma autónoma. Esto se puede lograr mediante la definición de una ruta (conjuntos de latitud y longitud) que el robot recorrerá utilizando su GPS incorporado; indicando un punto que será orbitado manteniéndolo siempre en foco; o incluso seguir a personas o vehículos en movimiento (DJI, 2015).

Esto implica que el Phantom 3 no solo es útil como un juguete o medio de entretenimiento, sino que es una herramienta ideal para producciones de películas con un presupuesto acotado.

Aunque el objetivo original de este dispositivo es ser utilizado por una persona con su control remoto, hay una gran variedad de funcionalidades (Liu, 2015) para las que se lo puede utilizar si se construye software que pueda controlar al dron en forma autónoma. Algunas de las aplicaciones prácticas que se le pueden dar incluyen la capacidad de patrullar un área, en busca de intrusos, conteo automático de ganado, e incluso análisis del estado de cosechas.

Pero para poder construir software que pueda tomar decisiones significativas es necesario acceder a la información que el dron dispone, y controlarlo en consecuencia, lo cual resulta ser más complejo de lo que debería.

Al día de hoy no existe forma de conectar directamente este dron a una computadora. Si bien existe para la versión 2 del Phantom un driver de comunicaciones del control remoto, el Phantom 3 no es compatible con este driver, lo cual pone un alto a cualquier investigación o desarrollo que se desee realizar con el mismo.

La empresa DJI solía comercializar una antena que, una vez conectada via USB a la computadora, permitía enviar mensajes al Dron utilizando comunicación serial. Sin embargo, el protocolo de comunicación que utilizaba era cerrado y no existe una especificación del mismo, obligando a realizar ingeniería inversa

para poder utilizarlo. Asimismo, desde el 12 de diciembre de 2015 Google dejó de soportar Google Earth, software sobre el cual estaba basada la tecnología de la Ground Station, por lo que DJI discontinuó este producto ¹

Líneas de Investigación, Desarrollo e Innovación

Este desarrollo en particular es un precursor para el futuro trabajo en Agentes y Sistemas Inteligentes.

Los componentes desarrollados se centran fuertemente en la línea de Procesamiento de señales y Sistemas de Tiempo Real

Resultados y Objetivos

El objetivo final de este desarrollo es generar todas las herramientas de software y hardware necesarias para poder controlar la Phantom 3 desde una computadora, accediendo a la información de los sensores y la cámara del mismo.

Dentro de este desarrollo se priorizarán aquellas opciones que simplifiquen la instalación y el uso del mismo para facilitar la distribución de esta tecnología de control.

Aunque no hay especificación abierta para controlar al dron desde la computadora, existe una API para dispositivos móviles (android, ios) que tiene todas las capacidades que podríamos querer para acceder y controlar al Phantom 3. Esto es de interés ya que esta API resuelve toda la comunicación con el control remoto, permitiendo enviar mensajes al dron y acceder a toda la información del dispositivo.

1. <http://forum.dji.com/thread-19865-1-1.html>

Se descartaron las librerías referentes a iOS ya que requieren permisos extra, pagar licencias de desarrollo de Apple, y no nos generaban ningun beneficio en particular.

Para el caso de Android, esta API cuenta con un conjunto de clases en Java que deben ejecutarse dentro de un ambiente de dicho sistema operativo. Por lo tanto, para poder utilizarlas se necesitó un sistema Android en funcionamiento. Esto se puede lograr con el uso de un dispositivo físico, o un emulador en una computadora.

La opción del emulador fue rápidamente descartada por tres grandes razones. En primer lugar, instalar y montar dicho emulador, y garantizar el acceso a los puertos usb reales a través de los cuales

necesaria para tanto controlar al Phantom 3 en vuelo como para evitar posibles accidentes.

Por último, la extensa presencia de dispositivos Android en el mercado Argentino aptos para este propósito inclinó la decisión hacia esta última opción.

Utilizando la API para dispositivos móviles entonces, se puede reemplazar la necesidad de tener un dispositivo físico como la DJI Ground Station utilizando un smartphone en su lugar. Dado que el drone puede comunicarse con un smartphone pero no directamente con una computadora, la solución propuesta es la utilización de un smartphone que oficie de proxy entre la computadora y el drone, como puede observarse en la Fig. 1



Fig. 1 Diagrama general de la solución

se conecta al control remoto de DJI es innecesariamente complejo. Aumentar la complejidad y los costos de implementación es algo a evitar.

En segundo lugar, el emulador tiende a tener tiempos de ejecución considerablemente más lentos que un dispositivo físico. Esto puede crear problemas donde los mensajes de control que deben pasar por el teléfono antes de llegar al dron sean demasiado lentos, y se pierda la capacidad de respuesta que es

Diagrama general de la solución.

Se generó una aplicación para android que una vez iniciada establece la conexión con el dron, pudiendo así acceder a la información de navegación, el streaming de video, y la capacidad de controlar al mismo. Esta conexión a su vez requiere que se validen permisos de control del dron contra uno de los servidores de DJI utilizando una identificación única de la aplicación.

Cuando esta validación se completa, la aplicación puede oficiarse como nexo entre la computadora y el Phantom 3.

La aplicación entonces abre un puerto TCP/IP al que puede conectarse la computadora para comunicarse. Los mensajes que la computadora envía a través de este puerto serán redirigidos al dron, y las respuestas del dron hacia la aplicación serán igualmente redirigidas hacia la computadora. De esta forma, el teléfono se convierte en un proxy de comunicaciones sin ninguna lógica aparte de la redirección de mensajes.

Resultados Obtenidos

La primera implementación fue sumamente simple. Una vez establecida la comunicación al servidor se envían ciertos caracteres al dispositivo móvil, cada uno de los cuales representa una dirección hacia la cual el Phantom 3 debe moverse. Si bien este método de control carece de precisión, fue suficiente para controlar al dron desde la computadora utilizando el software desarrollado.

Este simple protocolo de comunicación permitió la construcción del primer experimento. Se conectó un Kinect a la computadora, cámara desarrollada por Microsoft que permite obtener la posición en 3 dimensiones de las articulaciones de una persona, particularmente de las manos (Zhang, 2012). Para el experimento, se simuló la palanca de control de un avión de ala fija, ubicado en las manos del usuario, y se envió la información correspondiente al software de control en el smartphone. Esto permitió controlar al Phantom 3 de DJI utilizando gestos, cosa que no es posible con el esquema de control originalmente planteado por el fabricante.

El próximo paso en el desarrollo es mejorar el modelo de comunicación para

garantizar la disponibilidad completa de la información de navegación en la computadora. La primera implementación del software ubicado en el dispositivo móvil tiene acceso a dicha información, pero no la envía a la computadora.

Proveer toda la información disponible del dron al servidor garantiza que el mismo pueda tomar decisiones respecto a la posición, orientación, y estado (si se encuentra en el aire o no) del Phantom 3. Una vez conseguido esto se puede construir un sistema que haga al dron recorrer caminos, y aterrizar por sí mismo.

El último paso necesario es permitir la transmisión del video capturado por el Phantom. El dron envía actualmente el video al dispositivo Android usando el formato H.264 (Sullivan, 2012) uno de los más utilizados para transmitir video en la actualidad. Este formato se usa en los discos Blu-Ray, y en varios servicios de streaming como YouTube y Vimeo. Esta tarea en particular requiere establecer una segunda conexión entre la computadora y el teléfono, para garantizar que el envío y la recepción del video no interfieran con la información de control. Luego sólo queda transmitir cada bloque de H.264 recibido por el dispositivo móvil al servidor. De construirse esta segunda integración se podrían utilizar varias técnicas de análisis de imágenes para determinar el movimiento del dron, ofreciendo la capacidad de construir realmente un sistema de vuelo completamente autónomo y de un costo relativamente bajo.

Por otra parte, todos estos desarrollos tienen como objetivo generar un conjunto de librerías que simplifiquen el control del dron. De esta manera podrá ser programado por alumnos de la carrera de

Ingeniería en sistemas en diversas materias, con un nivel de complejidad definible según el grado de avance en la carrera. Utilizar un drone para estudiar aspectos de programación, control, robótica, procesamiento de imágenes u otros es altamente motivador para los estudiantes.

efficiency video coding (HEVC) standard. En *Circuits and Systems for Video Technology, IEEE Transactions on*, 22(12), 1649-1668.

Formación de Recursos Humanos

El equipo de trabajo está conformado por un investigador adjunto del Centro de Altos Estudios en Tecnología Informática (CAETI) quien ejerce el rol de director del proyecto, dos doctorandos, y un ayudante alumno de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana.

Referencias

Kakaes, K. (2015). What drones can do and how they can do it. En *Drones and aerial observation: new technologies for property rights, human rights, and global development*, 9.

Liu, Z., Li, Z., Liu, B., Fu, X., Ioannis, R., & Ren, K. (2015, June). Rise of Mini-Drones: Applications and Issues. En *Proceedings of the 2015 Workshop on Privacy-Aware Mobile Computing* (pp. 7-12). ACM.

DJI (2015, October). Phantom 3 Professional User Manual. En https://dl.djicdn.com/downloads/phantom_3/en/Phantom_3_Professional_User_Manual_V1.6.pdf.DJI

Zhang, Z. (2012). Microsoft kinect sensor and its effect. En *MultiMedia, IEEE*,19(2), 4-10.

Sullivan, G. J., Ohm, J. R., Han, W. J., & Wiegand, T. (2012). Overview of the high

Avances en robótica situada aplica a la navegaciones autónoma de cuadricópteros

Emiliano Lorusso, Diego Avila; Sofía Fasce, Gustavo Pereira, Norberto Mazza, Ierache Jorge

ISIER, Facultad de Informática Ciencias de la Comunicación y Técnicas
Especiales Universidad de Morón
Cabildo 134 Morón, Argentina
jierache@unimoron.edu.ar

Resumen

En este artículo se presentan la línea de investigación relacionada con el vuelo autónomo de un cuadricóptero, empleando un sistema de visión externa en el contexto de la robótica situada. En este orden se utiliza el mencionado sistema de visión para capturar el ambiente de actuación, conformado por cuatro puntos o checkpoint que el drone debe sobrevolar en forma autónoma.

Palabras claves: Robótica Situada, Robots Autónomos, Sistema de Visión, Navegación, Cuadricoptero, Drone

Contexto

El proyecto de desarrolla en el marco de investigación de grado, se radica en el Instituto de Sistema Inteligentes y Enseñanza experimental de la robótica de la FICCTE UM, en cooperación con la cátedra de Sistemas Inteligentes.

Introducción

El uso de drones para realizar diferentes tipos de tareas es de especial interés en la en la robótica autónoma. Se requiere una cantidad de parámetros que se deben tener en cuenta a la hora de moverse por el ambiente, así como también sus grados de libertad. Inicialmente se desarrolló un ambiente de robótica situado, en este campo se

encuentran distintas aplicaciones y competencias, como lo es el fútbol de robots, correspondiente a categorías físicas que reciben el ambiente del campo de juego desde una cámara suspendida. Un ejemplo que se puede mencionar es la categoría F-180 de Robocup [1]. La investigación plantea como problema la generación de un ambiente de robótica situada utilizando vehículos aéreos (dornes), considerando el control autónomo de los mismos a partir de información del ambiente provista por una cámara suspendida y la utilización de diferentes sensores disponibles en el dorne.

El modelo del ambiente de navegación utilizado en las pruebas de vuelo se representa en la Fig. 1. Se observa un campo de vuelo de 3 metros por 2 metros, definido por el ángulo de visión de la cámara, y ésta colocada en el centro del campo a 4 metros de altura. Dentro del campo, se situaron cuatro objetivos a alcanzar por el vehículo, que serán identificados a través del análisis de la imagen capturada por la cámara, de colores diferentes e irrepetibles. El checkpoint "A" de color amarillo; el checkpoint "B" de color verde; el checkpoint "C" de color naranja; y el checkpoint "D" de color negro. Todos los objetos identificables, son únicos dentro del ambiente para lograr un correcto

reconocimiento de los mismos.

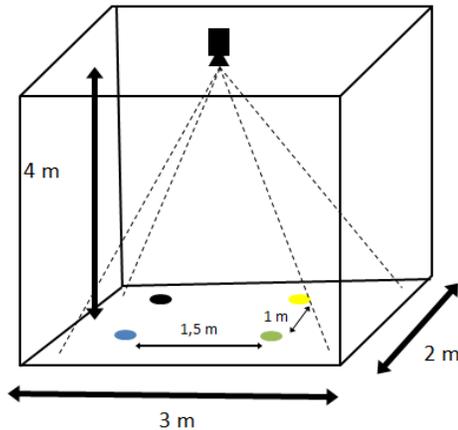


Fig. 1. Ambiente utilizado.

El control del dron se realiza mediante el análisis del video proveniente de la cámara suspendida, corrigiendo la trayectoria del mismo para mantenerlo estable. Se utilizó inicialmente un robot desarrollado por un grupo originario de Suecia, llamado CrazyFlie [2], que brindaba todas las herramientas necesarias para desarrollar y experimentar. Dentro de los parámetros que brinda el robot se encuentran: los ángulos de cabeceo (pitch), balanceo (roll) y guiñada (yaw). El desarrollo se hizo en el lenguaje C++, utilizando una librería llamada libcflie [3] necesaria para conectarse con el cuadricóptero, OpenCV [4] y un entorno Linux (Ubuntu 12.04). En la Fig. 2 se muestra una imagen del dron con un parche de color para su identificación.



Fig. 2. Cuadricóptero con identificación de color

Actualmente se trabaja con un robot ARDrone, de la empresa Parrot [5]. En la Fig. 3, se muestra la navegación del mismo sobre los checkpoint



Fig. 3 ARDrone ejecutando el recorrido correspondiente

El software desarrollado se encarga de controlar el dron y analizar el video, crea dos hilos de ejecución que comparten una porción de memoria permitiendo enviar mensajes entre ellos. Cada uno tiene un comportamiento diferente: el primero, se encarga de realizar el análisis del video proveniente de la cámara suspendida, y el segundo, se encarga de interpretar los datos obtenidos en el primer hilo y enviar las señales necesarias al dron para alcanzar el checkpoint. En la Fig. 4 se observa un diagrama general del software desarrollado. Se pueden ver los dos hilos de ejecución, correspondientes al video y al control del dron, y los datos compartidos a través de los cuales se envían información. Además, se puede ver cómo estos hilos se comunican con la cámara de video, la interfaz gráfica, y el dron respectivamente. Por otro lado, se desarrolló una interfaz que permite visualizar el análisis del video proveniente de la cámara y los valores resultantes luego de aplicar el algoritmo PID [6], a las variables cabeceo y balanceo del cuadricóptero.

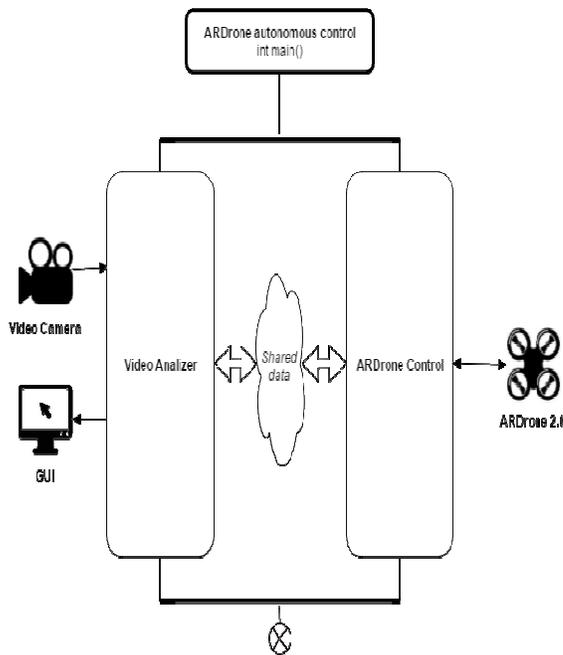


Fig. 4 Diagrama general del software desarrollado.

Líneas de Investigación, Desarrollo e Innovación

Las líneas de investigación se concentran en:

Sistemas de Visión para el desarrollo de la robótica situada aplicada a drones (cuadricópteros).

Implementación de sistemas de visión en tiempo real utilizando algoritmos de reconocimientos de imágenes para la detección de elementos situados en el ambiente. De esta manera se tiene total conocimiento de las posiciones de los objetos, pudiendo tomar decisiones sobre ellos.

Mecanismos de Navegación Autónomos de cuadricópteros en el contexto de robótica situada.

Se concentra el trabajo en el cálculo de los recorridos a realizar por el cuadricóptero [7], [8]; y la utilización de algoritmos como lo es el PID

(Proportional, Integral, Derivative), para la corrección de los errores propios del traslado del vehículo, a fin de lograr un vuelo autónomo, fluido y estable dentro del ambiente.

Mecanismos de Navegación Híbridos de cuadricópteros en el contexto de la robótica situada con el empleo de sensor de velocidad.

Esta línea se concentra en el desarrollo de navegación híbrida [9] basada en la utilización del sistema de visión externa junto con los sensores provistos por el dron (sensor de velocidad), logrando así tener mayor control sobre las variables que afectan al vuelo.

Ambiente de robótica situada aplicada a drones.

Esta línea de trabajo se concentra en el desarrollo de una herramienta que brinde el ambiente, la posición del dron, la posición de los checkpoint, permitiendo a un tercero incluir sus mecanismos de navegación. De esta forma se puede generar un ambiente de competencia de robótica situada aplicada a drones.

Resultados y Objetivos

Inicialmente se realizaron pruebas de vuelo donde el cuadricóptero debía sobrevolar los checkpoint establecidos en un circuito. Se documentó el tiempo que le llevaba al dron completar el circuito, y estos se compararon con el tiempo ideal en el que debía completar el mismo.

Los resultados para las pruebas mencionadas de navegación en relación a una navegación de vuelo ideal dieron una diferencia del orden del 14 % en función de la comparación de tiempos

de navegación para que el dron efectúe la navegación cruzando los cuatro checkpoint del ambiente de prueba. [7], [8] y un vuelo ideal considerando su velocidad constante. El principal desvío sucede a la hora del despegue, entre que el robot se desprende del suelo y acondiciona su vuelo hacia el primer checkpoint.

Luego cuando se comenzó a experimentar con el cuadricóptero ARDrone se realizaron dos conjuntos de pruebas, en ambos casos el dron debía cumplir con un circuito establecido. El primer conjunto de pruebas se basaba en un modelo de vuelo exclusivamente de visión, ya que la misma se utilizaba tanto como para localizar los objetos dentro del ambiente, como para calcular la velocidad de vuelo. Al segundo conjunto de pruebas lo denominamos híbrido, ya que en este caso, para el obtener la velocidad se utilizaba el sensor propio del cuadricóptero. Inicialmente este modelo híbrido no se podía realizar ya que el dron CrazyFlie no tenía este tipo de sensores.

En la utilización de vehículos aéreos en el contexto de la robótica situada y autónoma, es imprescindible la utilización de una mecánica de vuelo que brinde cierta fluidez para moverse dentro del ambiente, y la mayor cantidad de datos respecto del estado del vehículo y del contexto. Estas dos características fundamentales, se consiguen no sólo realizando un análisis del video proveniente de la cámara suspendida, sino también mediante la utilización de todos los sensores de los que disponga el robot a utilizar. Se demuestra experimentalmente en la duración del tiempo de vuelo del dron, para la realización de los dos patrones de vuelo una diferencia promedio del 20% entre un modelo híbrido al aplicar la velocidad del dron y la determinación de posición del mismo

por el sistema de visión, frente a un modelo centrado en la visión desde el cual se determina además de la posición la velocidad. Esto es así ya que los datos obtenidos a través de los sensores del cuadricóptero en comparación con los obtenidos mediante el análisis del video son más precisos.

En el contexto de las futuras líneas de trabajo se explora la posibilidad de realizar la conexión del cuadricóptero con una interface cerebro-computadora (BMI) EmotiveEpoC. [10], [11]. Teniendo esta conexión se podrá controlar los movimientos del cuadricóptero por medio de bioseñales, enviando al mismo distintos comandos de control. Fusionando estas dos tecnologías se logrará realizar diversas acciones como lo son, seleccionar los checkpoints a sobrevolar según estímulos de bioseñales o mover el cuadricóptero hacia la dirección pensada.

Formación de Recursos Humanos

El equipo de trabajo se conforma de un investigador formado, dos investigadores en formación y tres investigadores estudiantes de la carrera de ingeniería informática. Se relacionan con las líneas de investigación tres tesis próximas a defender

Referencias

1. Robocup, http://wiki.robocup.org/wiki/Small_Size_League
2. Bitcraze AB, <http://www.bitcraze.se>
3. Crazyflie Nano C++ Client Library, <https://github.com/fairlight1337/libcflie>
4. OpenCV, <http://opencv.org/>

5. ARDrone 2.0 de Parrot,
<http://ardrone2.parrot.com/>

6. Jinghua Zhong. PID Controller Tuning: A short tutorial. Purdue University, 2006.

<http://saba.kntu.ac.ir/eecd/pcl/download/PIDtutorial.pdf>

7. Ávila, D. Lorusso, E. Fasce, S. Ierache, J.: Robótica situada aplicada al control de vehículos aéreos. CACIC 2014. <http://hdl.handle.net/10915/42255>

8. Ávila, Diego, Lorusso, Emiliano, Fasce, Sofia, Pereira, Gustavo, Ierache, Jorge Salvador. Experiencias en el control de un Drone en el contexto de la robótica situada. Jornadas Argentinas de Tecnología, Innovación y Creatividad JATIC Mar del Plata Noviembre 2015 ISBN 978-987-23963-2-9.

9. Sofia Fasce, Diego E. Ávila, Emiliano Lorusso, Gustavo Pereira, Jorge Ierache “Autonomous control of a Drone in the context of Situated Robotics”. Robot Intelligence Technology and Applications 4 Serie Advances in Intelligent Systems and Computing of Springer. (En prensa)

10 Navigation Control of a Robot from a remote location via the Internet using Brain-Machine Interface Ierache Jorge, Pereira Gustavo, Iriibarren Juan; Robot Intelligence Technology and Applications 2. : Springer. 2014. p - 297-310. ISBN: 978-3-319-05581

11. Ierache Jorge, Pereira Gustavo, Iriibarren Juan, Sattolo Iris, “Robot Control on the Basis of Bio-electrical Signals”, Robot Intelligence Technology and Applications, Series Advances in Intelligent and Soft Computing of Springer. ISBN: 978-3-642-37373-2 (Print) 978-3-642-37374-9 (Online), pp 337-346, Volumen 208, 2013.

Sistemas de Tiempo Real Mixtos: Planificación en Sistemas Operativos de Tiempo Real Bajo Plataformas de Desarrollo Concretas

José M. Urriza, Mariano Ferrari, Javier D. Orozco, Ricardo Cayssials,
Francisco E. Páez, Gabriela Olguín, Lucas Schorb, Sebastián Lucas,
Rodrigo Tolosa

Departamento de Informática – Facultad de Ingeniería – Universidad Nacional
de la Patagonia San Juan Bosco – Sede Puerto Madryn
Bvln. Brown 3051, Puerto Madryn, Chubut
josemurriza@gmail.com

Resumen

Actualmente, en numerosos centros de investigación alrededor del mundo, este tema es investigado de manera intensiva, dado que es fundamental proveer de herramientas y soluciones, que permitan implementaciones tecnológicas que maximicen las prestaciones de los recursos computacionales, alcanzando así el menor costo posible y maximizando el ciclo de vida del producto final. Por otro lado, un diseño apropiado permite cumplir con los requerimientos establecidos por las aplicaciones, en conjunto con otros requerimientos adicionales como: el correcto manejo de la calidad de servicio para la ejecución de conjuntos de tareas mixtas y/o heterogéneas, el ahorro de energía, la robustez y la tolerancia a las fallas, entre otras.

Por otro lado, es importante encontrar un balance apropiado entre una implementación sencilla, robustez y eficiencia, dado que posibilita extender el ciclo de vida del software de tiempo real y mejorar las herramientas para el desarrollo de nuevas tecnologías, validación de la especificación, diseño, optimización y testeo.

Este proyecto plantea como objetivo general, el desarrollo de técnicas de modelado, diseño, análisis, optimización y testeo de sistemas embebidos con restricciones de tiempo y requerimientos mixtos, para diferentes plataformas de software y hardware, con un apropiado balance entre prestaciones, rendimiento y eficiencia.

Palabras clave: Sistemas de Tiempo Real Mixtos, Sistemas Embebidos, Sistemas Operativos de Tiempo Real

Contexto

El Proyecto “Sistemas de Tiempo Real Mixtos: Planificación en Sistemas Operativos de Tiempo Real Bajo Plataformas de Desarrollo Concretas” se encuentra inserto en la líneas de investigación de Sistemas de Tiempo Real (*STR*), Sistemas Embebidos (*SE*) y Sistemas Operativos de Tiempo Real (*SOTR*). En particular se busca la implementación sobre plataformas concretas de desarrollo como: Arduino, Arduino Due, mbed, CIAA, Raspberry Pi, Raspberry Pi B+, PCduino, las cuales el grupo de investigación en *STR* de la

UNPSJB sede Puerto Madryn posee en número.

El proyecto pretende desarrollar nuevas investigaciones en estas temáticas, mejorar las existentes, e implementarlas en algunas de las plataformas antes mencionadas.

Además se espera una activa participación de los doctores Orozco y Cayssials, pertenecientes al Departamento de Ingeniería Eléctrica y Computadoras de la UNS, y de larga trayectoria en esta temática, como miembros asesores.

El proyecto está acreditado por la Secretaria de Ciencia y Técnica de la UNPSJB, la cual también lo financia.

Introducción

Numerosos *SE* son utilizados en aplicaciones que poseen restricciones temporales, en la industria o situaciones cotidianas. Es por ello que resulta frecuente que posean un *SOTR*. Estos en muchos casos, suelen no ser eficientes en varios aspectos. Consecuentemente, el presente proyecto, investiga las técnicas, algoritmos y modelos actuales factibles de optimización para mejorar la eficiencia de los mismos.

En los Sistemas Embebidos de Tiempo Real (*SETR*), es complejo alcanzar altos niveles de eficiencia en el manejo de recursos, hecho que usualmente resulta en un compromiso de diseño. Es deseable por costo y consumo, contar con una arquitectura e implementación de hardware suficiente, pero no excesivamente sobredimensionada, la cual incrementaría costos, y generaría otro tipo de requerimientos que muchas veces va en contra de los objetivos de diseño planteados. Por ejemplo, sobredimensionar el procesador trae acarreado un incremento en el consumo de energía, lo que genera mayor disipación de calor y, si el sistema es

móvil, mayor consumo de batería, etc. Este punto es un desafío ya que existen numerosos desarrollos teóricos que, en la práctica, no están disponibles para resolver ciertas clases de aplicaciones.

Estos temas son investigados desde diversos puntos. Primero, es encontrar herramientas de diseño y testeo que permitan minimizar los recursos empleados por el sistema. Segundo, es como utilizar el sobrante en potencia de cómputo para ejecutar tareas mixtas, ahorrar energía, etc.

Además, en la actualidad, tan solo el 2% del total de los microprocesadores fabricados, corresponden a computadoras personales (grandes procesadores). Por lo tanto, el 98% restante son utilizados en dispositivos de propósito dedicado o *SE* (como teléfonos, microondas, etc.) con lo cual es necesario adecuar las herramientas de diseño, y desarrollar las que fueren necesarias, para abordar esta problemática particular, para darle a la industria soluciones eficientes. Esto incluye herramientas de análisis y clasificación de requerimientos, modelado, diseño, validación y verificación.

Lo expuesto revela una necesidad disciplinar, que es la principal motivación del presente proyecto: desarrollar nuevas técnicas analíticas en el área de la planificación de sistemas tiempo real y en la medida de lo posible, su aplicación experimental a fin de evaluar su rendimiento.

Como motivaciones socioeconómicas se pueden nombrar: las reducciones en los costos de producción, menor consumo de energía, miniaturización del dispositivo, menor contaminación ambiental debido al menor consumo, menores requerimientos de hardware, mejores prestaciones para los usuarios acompañada de reducción de costos de equipos, desarrollo de nuevas aplicaciones, mejor adaptabilidad de los

diseños a las crecientes reducciones del ciclo de vida del producto, etc.

Desde el punto de vista regional, se generará nuevos recursos humanos de alta capacidad. Esto permitirá mejorar las cadenas de valor, por el aporte de competencias distintivas en diferentes áreas de la producción industrial, desde la innovación de segmentos clásicos de la producción a la captación de nuevas áreas productivas alrededor del desarrollo de la electrónica y el software embebido. Esto beneficia en particular, satisfacer la demanda generada por la nueva generación de productos del área de las TICs. A tal fin, este proyecto incluye alumnos, futuros profesionales que se verán capacitados en esta disciplina.

Líneas de Investigación, Desarrollo e Innovación

Las líneas de investigación del proyecto “Sistemas de Tiempo Real Mixtos: Planificación en Sistemas Operativos de Tiempo Real Bajo Plataformas de Desarrollo Concretas” son: Sistemas de Tiempo Real (*STR*), Sistemas Embebidos (*SE*) y Sistemas Operativos de Tiempo Real (*SOTR*) y sus subramas de las mismas.

Resultados y Objetivos

Los objetivos pueden resumirse en los siguientes:

- Continuar la investigación sobre los *SOTR* más relevantes para sistemas embebidos (*FreeRTOS*, *RTOSosek*, *MarteOS*, etc.), en particular cómo implementan la administración de sus recursos, cumpliendo con las constricciones temporales impuestas por las tareas que conforman el sistema. También, seguir investigando las plataformas en las cuales estos *SOTR* ejecutan.

- Determinar la factibilidad y aplicabilidad de los métodos teóricos en los entornos prácticos estudiados. Proponer mejoras o nuevas técnicas y/o re-formulaciones a las técnicas existentes para el manejo de recursos, temporales y espaciales.
- Implementar y validar las técnicas y métodos propuestos sobre plataformas de desarrollo concretas, modificando los *kernels* de los *SOTR* mencionados o mediante tareas administradoras de recursos.
- Publicar las investigaciones realizadas. Además, fomentar y difundir las tareas de investigación en la Universidad con trabajos y cursos.
- Mejorar la formación de recursos humanos, brindando gran participación a los alumnos.

Los resultados del proyecto todavía no pueden evaluarse dado que el mismo comienza en octubre del presente año. Sin embargo en las referencias y bibliografía se puede observar la producción de proyectos actuales y pasados del grupo de investigación, todos en líneas de investigación relacionadas con el presente proyecto.

Como resultados del proyecto, además de los trabajos académicos y de divulgación científica, se espera mejorar la eficiencia de dispositivos de propósito dedicado desarrollando, mejorando e implementando nuevos métodos y técnicas de planificación dentro del núcleo de un *SOTR*.

En los proyectos que anteceden a este, y que son base para la ejecución del mismo, se ha construido un simulador de *STR*, un generador de *STR* para alimentar al simulador y se han publicado varios trabajos en congresos nacionales e internacionales, así como trabajos en

revistas ([1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]).

Se han implementado métodos de *Slack Stealing (SS)* en los *SOTR MarteOS (PC)* [20] y *FreeRTOS (Arduino, mbed, CIAA)* [21], para la administración del tiempo ocioso. La implementación sobre *FreeRTOS*, además, está diseñada para facilitar la programación de nuevos algoritmos de cálculo de *slack* en línea. Se desarrolló también un *framework* para implementar políticas de planificación a *nivel de usuario*, denominado *HST*, sobre *FreeRTOS* [22]. Mediante *HST* se han implementado políticas de planificación tales como *EDF*, *RM+SS* y *Dual Priority*, sin necesidad de alterar el *kernel* del *SOTR*. El código fuente de estos desarrollos está disponible en el sitio web del *Real Time Systems Group UNPSJB Sede Puerto Madryn*¹. Por medio de estas herramientas y desarrollos se facilitara la implementación y evaluación de cualquier nueva técnica resultado del trabajo en el presente proyecto.

Actualmente, se está implementado un método de *SS* en el *kit* Lego Mindstorm NXT 2.0 (*NXT OSEK*), cuya finalización se dará en el marco de este proyecto.

Formación de Recursos Humanos

En este proyecto de investigación participan 9 personas, de las cuales 6 son docentes, 4 de los mismos son docentes con el grado de doctor y 1 está en proceso de obtener su título de doctor antes de que finalice el presente proyecto, con una Beca Doctoral de CONICET. Además, los 3 alumnos son de la carrera Licenciatura en Informática de la Sede Puerto Madryn de la UNPSJB, los cuales realizan actividades dentro del grupo de investigación en STR y han comenzado

sus tesinas de grado que se espera que finalice en el marco de este proyecto

Referencias

- [1] B. Novelli, J. C. B. Leite, J. M. Urriza, and J. D. Orozco, "Regulagem Dinâmica de Voltagem em Sistemas de Tempo Real," in *XXXII Seminário Integrado de Software e Hardware (SBC 2005 SEMISH)*, Unisinos -Sao Leopoldo, Brazil, 2005.
- [2] J. M. Urriza, R. Cayssials, and J. D. Orozco, "Optimización del Cálculo del Tiempo Ocioso en Planificadores DVS con Tiempos de Ejecución Variables," in *XXXII Conferencia Latinoamericana de Informática, Clei 2006*, Santiago, Chile, 2006.
- [3] J. M. Urriza, R. Cayssials, J. D. Orozco, and J. C. B. Leite, "Modelo de Tareas para recuperacion de Slack para Aplicaciones en Sistemas Enbebidos con DVS," Dep. de Ing. Eléctrica y Computadoras, Universidad Nacional del Sur, Argentina., Bahía Blanca, Reporte Interno22 de Mayo 2005.
- [4] J. M. Urriza, B. Novelli, J. C. B. Leite, and O. Javier Dario, "Economia de energia em dispositivos móveis," in *VI Workshop de Comunicação sem Fio e Computação Móvel*, Fortaleza, CE, Brasil, 2004, pp. 48–56.
- [5] J. M. Urriza, R. Cayssials, and J. D. Orozco, "A Fast Slack Stealing Method for embedded Real-Time Systems," Dep. de Ing. Eléctrica y Computadoras, Universidad Nacional del Sur, Argentina., Bahía Blanca, Internal ReportMay 31 2005.
- [6] J. M. Urriza, J. D. Orozco, and R. Cayssials, "Fast Slack Stealing methods for Embedded Real Time Systems," in *26th IEEE International Real-Time Systems Symposium (RTSS 2005) - Work In Progress Session*, Miami, EEUU, 2005, pp. 12-16.
- [7] R. M. Santos, J. M. Urriza, J. Santos, and J. D. Orozco, "New methods for redistributing slack time: applications and comparative evaluations," *The Journal of Systems & Software*, vol. 70-2, pp. 115-128, 2004.
- [8] J. M. Urriza, R. Cayssials, and J. D. Orozco, "Optimización on-line de Sistemas de Tiempo Real con Computación Imprecisa Basados en Recompensas.," in *32 JAIIO AST2003*, Buenos Aires, Argentina, 2003.
- [9] J. M. Urriza, J. Santos, and J. D. Orozco, "Un Algoritmo para la Diagramación de Tareas No-Duras mediante el Cálculo del Slack Time Disponible en cada Instante," in *XXIX Conferencia Latinoamericana de Informática. CLEI*, Bolivia, La Paz, 2003.
- [10] J. M. Urriza, J. D. Orozco, R. Cayssials, and L. Schorb, "Reduced Computational Cost in the Calculation of Worst Case Response Time for Real Time Systems," *Journal of Computer Science & Technology*, vol. 9, pp. 72-81, 2009.
- [11] J. M. Urriza, J. D. Orozco, C. Buckle, and R. Cayssials, "Ahorro de Energía en Dispositivos con un SO de Tiempo Real que planifican en RM o

¹ <http://www.rtsg.unp.edu.ar>

- DM," in *Encuentro Chileno de Computacion*, Santiago, Chile, 2009.
- [12] J. M. Urriza, R. Cayssials, and E. Ferro, "Hardware Co-Processing Unit For Real-Time Scheduling Analsys," presented at the VI Southern Conference on Programmable Logic 2010 - Designer Forum 2010, Ipojuca Porto Galinhas Beach, Brazil, 2010.
- [13] J. M. Urriza, F. E. Paez, R. Cayssials, J. D. Orozco, and L. Schorb, "Low Cost Slack Stealing Method for RM/DM," *International Review in Computers and Software (IRECOS)*, vol. 5, pp. 660-667, 2010.
- [14] G. Olguín, L. Biscayart, and J. M. Urriza, "Generador de Conjuntos de Tareas para Simulación en Sistemas de Tiempo Real," presented at the JAIIO 39 - Jornadas de Informática Industrial (JII) 2010, Buenos Aires, 2010.
- [15] C. E. E. Buckle, J. M. Urriza, and F. E. Paez, "Transitando Hacia las Bases de Datos de Tiempo Real," in *JAIIO 39 - Jornadas de Informática Industrial (JII) 2010*, Buenos Aires, 2010, pp. 1901-1915.
- [16] L. Moreno, C. Geymonat, and J. M. Urriza, "Conceptos de Tiempo Real Aplicados a la Informática Industrial," presented at the CACIC 2011 - XVII Congreso Argentino de Ciencias de la Computación, La Plata, Arg., 2011.
- [17] F. E. Paez, J. M. Urriza, J. D. Orozco, and C. R. Buckle, "Un Modelo de Eventos Discretos para la Simulación de Sistemas de Tiempo Real," presented at the CACIC 2011 - XVII Congreso Argentino de Ciencias de la Computación, La Plata, Arg., 2011.
- [18] C. E. Buckle, J. M. Urriza, D. P. Barry, and F. E. Paez, "Tipo de Dato Abstracto para Sistemas de Bases de Datos de Tiempo Real," presented at the CACIC 2011 - XVII Congreso Argentino de Ciencias de la Computación, La Plata, Arg., 2011.
- [19] G. Olguin, L. Biscayart, and J. M. Urriza, "Generación de Tareas Periódicas y Aperiódicas para Simulación de Sistemas de Tiempo Real," presented at the JAIIO 40 - Jornadas de Informática Industrial 2011, Cordoba, Arg., 2011.
- [20] L. A. Díaz, F. E. Páez, J. M. Urriza, J. D. Orozco, and R. Cayssials, "Implementación de un Método de Slack Stealing en el Kernel de MaRTE OS," *43 JAIIO (Jornadas Argentinas de Informática)*, vol. 3º Simposio Argentino de Informatica Industrial (SII), pp. 13-24, Septiembre 2014.
- [21] F. Paez, J. M. Urriza, R. Cayssials, and J. D. Orozco, "Métodos de Slack Stealing en FreeRTOS," in *44 Jornadas Argentinas de Informática (JAIIO)*, Rosario, 2015.
- [22] F. E. Paez, J. M. Urriza, R. Cayssials, and J. D. Orozco, "FreeRTOS user mode scheduler for mixed critical systems," in *Embedded Systems (CASE), 2015 Sixth Argentine Conference on*, 2015, pp. 37-42.

Sistemas de Tiempo Real Distribuidos Robots y Microcontroladores

Fernando Romero, Mariano Méndez , Diego Encinas,
Santiago Medina, Armando De Giusti¹, Fernando G. Tinetti²

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata

{fromero, mmendez, dencinas, smedina, degiusti, fernando}@lidi.info.unlp.edu.ar

Resumen

El objetivo de esta línea de investigación es el estudio y desarrollo de sistemas de software que poseen restricciones temporales, como son los Sistemas de Tiempo Real (STR), atendiendo en especial los aspectos relacionados con planificadores y comunicaciones, realizando implementaciones de robots y en general sistemas de adquisición y control utilizando microcontroladores.

Contexto

Esta línea de Investigación forma parte del proyecto "Arquitecturas multiprocesador en HPC: software de base, métricas y aplicaciones" (11/F018) del Instituto de Investigación en Informática LIDI acreditado por la UNLP.

Palabras Claves: Tiempo Real, Sistemas Embebidos, Comunicaciones, Red de Sensores, Robots, Microcontroladores.

1. Introducción

Los Sistemas de Tiempo Real (STR) son sistemas que interactúan con su entorno

físico y deben satisfacer restricciones de tiempo de respuesta explícitas [14] [7] [15] [8] [23]. El requerimiento de satisfacer los plazos llega a ser tanto o más importante que la corrección lógica de la respuesta, que puede incluso relajarse en función de cumplir los plazos. Son en su mayoría Sistemas Embebidos, en cuanto controlan un sistema, por lo general físico, adquiriendo información de las variables del sistema controlado a través de sensores, detectores y realizando el control a través de distintos tipos de actuadores. Muchos de estos sistemas son móviles, con lo que el aspecto consumo se debe tener en cuenta. Esto requiere el uso de hardware con bajo consumo energético, tales como los microcontroladores. En los microcontroladores se tiene una menor capacidad de cómputo, y a la vez una sofisticada sección de entrada y salida. Asimismo, los SOTR (Sistemas Operativos de Tiempo Real) que utilizan están optimizados para lograr baja latencia en la elaboración de las respuestas frente a señales externas, por lo general a través de interrupciones de hardware. Parte de las aplicaciones actuales de estos sistemas pertenecen al

¹ Investigador CONICET

² Investigador CIC Prov. de Buenos Aires.

campo de la robótica, que combina el control con cierto grado de inteligencia [5] [14], creando sistemas que llevan a cabo tareas que ni las máquinas comunes ni los seres humanos pueden realizar.

La satisfacción de los requerimientos temporales de los plazos de estos sistemas implica una planificación rigurosa de la ejecución de tareas. Este es el punto central del diseño de aplicaciones de tiempo real. Es objetivo de esta línea de investigación abordar este tema, desde el punto de vista teórico y experimental. Esta planificación puede ser realizada en forma manual, durante el diseño del software, o utilizar herramientas automáticas para la misma, conocidas como planificadores, mayormente formando parte de los SOTR. El grado de complejidad de estos planificadores será acorde al grado de complejidad del sistema. Para este estudio se dispone de diversas plataformas de hardware y software para la construcción de sistemas reales, de diferente grado de complejidad, desde sistemas basados en microcontroladores simples sin SOTR, como Arduino hasta sistemas basados en PC, pasando por microcontroladores sofisticados capaces de soportar SOTR basados en Linux, tales como los Freescale [32] y CIAA [27]. Los sistemas más simples implican abordar el estudio del nuevo paradigma llamado “Internet of Things” (IoT) [22] [21] [3], en el cual cosas u objetos (sensores, actuadores, tabletas, teléfonos celulares, etc.) que se conectan en un esquema de identificación única son capaces de interactuar unos con otros y cooperar.

En las diferentes implementaciones que se llevan a cabo en esta línea de investigación se trabaja de manera experimental con mini-robots, plataformas de cómputo (PC, Arduino, Intel Galileo, Freescale, CIAA, etc.), con diferentes sistemas

operativos (Linux RT-Preempt, Freertos, MQX, OSEK-OS, etc.) [20] [10] y sistemas de comunicación basados en los protocolos I2C, CANBUS [29] [30] [31] [32] [33] y MODBUS [28].

Tratándose en algunos casos de sistemas distribuidos, se experimentan sistemas de microcontroladores [11] interconectados con PC, en forma cableada e inalámbrica (radiofrecuencia, Bluetooth, WiFi). Esto implica el estudio de protocolos de comunicación diseñados para este tipo de aplicaciones, tales como CANBUS y MODBUS. Con respecto a los micro-robots, se utilizan algunos de diseño y fabricación propia como también los sofisticados robots Khepera [24] [25] [26]. Una característica de los SOTR es tener un alto grado de fiabilidad, por lo que el estudio de la detección y control de condiciones de falla es un tema importante [2] [17].

2. Líneas de Investigación y Desarrollo

Se plantean como temas de estudio:

- Verificación y validación del hardware, donde pueden encontrarse detalles o resultados de simulaciones que impliquen modificaciones [9] [10].
- Construcción de red de sensores en la plataforma de microcontroladores interactuando con una PC, utilizando MODBUS y CANBUS como protocolo de comunicaciones [1] [18].
- Diseño y construcción de vehículos no tripulados y sistema de navegación [12].
- Estudio de diferentes plataformas y combinaciones de las mismas.
- Odometría con robots móviles, en particular con los robots Khepera [6] [16] [4].

3. Resultados Esperados y Objetivos

De acuerdo con las tareas desarrolladas y a desarrollar, los resultados se enfocarán en varias direcciones relacionadas con los sistemas de tiempo real:

- Construcción y uso de robots tipo vehículo autónomo para el estudio de Odometría.
- Medición de consumo de diferentes tipos de microcontroladores bajo distintas condiciones de uso.
- Estudio de distintas plataformas de hardware: Arduino, Intel Galileo, CIAA, FreeScale, Rasperry Pi, entre otras.
- Construcción y estudio de dos redes de Sensores, empleando rf para las comunicaciones en una y los protocolos CANBUS y MODBUS en otras.

4. Formación de Recursos Humanos

En base a estos temas se están desarrollando trabajos de varios alumnos encuadrados en el marco de la Convocatoria a Proyectos de Desarrollo e Innovación de la Fac. de Informática de la UNLP como también de PPS (Práctica Profesional Supervisada) con la que concluyen sus estudios los alumnos de Ingeniería en Computación.

5. Referencias

[1] Akyildiz, Ian F., Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. "Wireless sensor networks: a survey." *Computer networks* 38, no. 4 (2002): 393-422.

[2] Andersen, B. L. "Method of detecting systemic fault conditions in an intelligent electronic device." U.S. Patent 6,434,715, issued August 13, 2002.

[3] Atzori, L., A. Iera, G. Morabito. "The internet of things: A survey." *Computer networks* 54, no. 15 (2010): 2787-2805.

[4] Azizi, F., N. Houshangi. "Mobile robot position determination using data from gyro and odometry." In *Electrical and Computer Engineering*, 2004. Canadian Conference on, vol. 2, pp. 719-722. IEEE, 2004.

[5] Bekey, George A. *Robotics: state of the art and future challenges*. Imperial College Press, 2008.

[6] Borenstein, Johann, Liqiang Feng, "Gyrodometry: A new method for combining data from gyros and odometry in mobile robots." In *Robotics and Automation*, 1996. Proceedings., 1996 IEEE International Conference on, vol. 1, pp. 423-428. IEEE, 1996.

[7] Burns, A, A, Wellings. "Real-Time Systems and Programming Languages: Ada, Real-Time Java and C/Real-Time POSIX", Addison-Wesley Educational Publishers Inc., 2009.

[8] Buttazzo, G. C., "Hard RealTime Computing Systems", Third edition, Springer, 2011.

[9] Eickhoff, J., *Simulating Spacecraft Systems*, Springer, 2009.

[10] "FreeRTOS - market leading RTOS (real time operating system) for embedded systems supporting 34 microcontroller architectures". <http://www.freertos.org/>.

[11] Jenkins, T., I. Bogost. "Designing for the internet of things: prototyping material interactions." In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pp. 731-740. ACM, 2014.

[12] Jones, J. L., A. M. Flynn, Bruce A. Seiger. *Mobile robots: inspiration to*

- implementation. Vol. 2. Wellesley MA: AK peters, 1999.
- [13] Kopetz. H., "Real-Time Systems, Design Principles for Distributed Embedded Applications". Second Edition. Springer. 2011.
- [14] Krishna, C. Mani. Real-Time Systems. John Wiley & Sons, Inc., 1999.
- [15] Liu, J. W. S. Liu, "Real Time Systems", Integre Technical Publishing Co., Inc., 2000
- [16] Rekleitis, I. M., G. Dudek, E. E. Milios. "Multi-robot exploration of an unknown environment, efficiently reducing the odometry error." In International Joint Conference on Artificial Intelligence, vol. 15, pp. 1340-1345. Lawrence Erlbaum Associates Ltd, 1997.
- [17] Rohani, A., H. R. Zarandi. "An analysis of fault effects and propagations in AVR microcontroller ATmega103 (L)." In Availability, Reliability and Security, 2009. ARES'09. International Conference on, pp. 166-172. IEEE, 2009.
- [18] Savvides, A., M. B. Srivastava. "A distributed computation platform for wireless embedded sensing." In Computer Design: VLSI in Computers and Processors, 2002. Proceedings. 2002 IEEE International Conference on, pp. 220-225. IEEE, 2002.
- [19] Silberschatz, A., P. B. Galvin, G. Gagne, Operating System Concepts, 8th Edition, ISBN : 978-0-470-12872-5, Wiley, 2009.
- [20] The Cheddar project: A free real time scheduling analyzer, <http://beru.univbrest.fr/~singhoff/cheddar/>
- [21] Weber, Rolf H., and Romana Weber. Internet of Things. New York: Springer, 2010.
- [22] Xia, Feng, L. T. Yang, L. Wang, and Alexey Vinel. "Internet of things." International Journal of Communication Systems 25, no. 9 (2012): 1101.
- [23] PHILLIP A. LAPLANTE, SEPPO J. OVASKA. REAL-TIME SYSTEMS DESIGN AND ANALYSIS Tools for the Practitioner Fourth Edition. A JOHN WILEY & SONS, INC., PUBLICATION. IEEE PRESS. 2012.
- [24] <http://ftp.k-team.com/khepera/documentation/Kh2UserManual.pdf>
- [25] <http://ftp.k-team.com/khepera/documentation/Kh2ProgrammingManual.pdf>
- [26] <http://ftp.k-team.com/khepera/documentation/Kh2IRAN.pdf>
- [27] <http://www.proyecto-ciaa.com.ar/devwiki/doku.php?id=desarrollo:edu-ciaa:edu-ciaa-nxp>
- [28] <http://www.tolaemon.com/docs/modbus.htm>
- [29] http://www.bosch-semiconductors.de/en/ubk_semiconductors/ip_modules_3/produkttable_ip_modules/can_literature_1/can_literature.html
- [30] <http://www.can-cia.de/can-knowledge/can/can-fd/>
- [31] http://www.iso.org/iso/catalogue_detail.htm?csnumber=59165
- [32] http://www.freescale.com/files/microcontrollers/doc/app_note/AN1798.pdf
- [33] http://www.can-cia.org/W_Scanner_Auto_OBD2_OBD1_2x2_E_OBD_20_PIN.html

Sintonía de Controladores Inteligentes Mediante Estrategia Híbrida Fuzzy-PSO

Miguel A. Azar; Sergio L. Martínez; Enrique E. Tarifa
Samuel Franco Domínguez; Jorge J. Gutiérrez

Facultad de Ingeniería / Universidad Nacional de Jujuy / CONICET
Ítalo Palanca N° 10 / S. S. de Jujuy / Provincia de Jujuy / Tel. 0388-4221591
maazar@fi.unju.edu.ar; smartinez@fi.unju.edu.ar; eetarifa@fi.unju.edu.ar
sfrancodominguez@gmail.com; jorgejgutierrez@outlook.com

Resumen

El presente trabajo se desarrolla en el contexto del proyecto de investigación denominado “Diseño y Simulación de Controladores Inteligentes Aplicados a Procesos Industriales”. El trabajo está orientado en la aplicación de técnicas no convencionales para el ajuste y sintonía de los controladores inteligentes, implementando la técnica de Optimización por Enjambre de Partículas (PSO) en tándem, sobre un controlador de lógica fuzzy cuyos parámetros son convenientemente modificados por el algoritmo PSO.

En el diseño de controladores convencionales aplicados a procesos productivos o industriales, existe una serie de técnicas para lograr la sintonía adecuada. Por otra parte, el desarrollo de controladores inteligentes requiere significativos esfuerzos de investigación interdisciplinarios para integrar conceptos y métodos de áreas tales como control, identificación, estimación, además de otras áreas tales como teoría de la comunicación, ciencias de la computación, inteligencia artificial e investigación operativa. Debido a su estructura particular, el ajuste o sintonía de sistemas controladores no convencionales, basados en inteligencia artificial, puede realizarse mediante la modificación de los parámetros intrínsecos del controlador.

Palabras clave: Controladores inteligentes, sintonización, lógica fuzzy, procesos industriales, PSO.

Contexto

El trabajo de ajuste de controladores inteligentes mediante Fuzzy-PSO que se describe, se encuentra enmarcado en el proyecto “Diseño y Simulación de Controladores Inteligentes Aplicados a Procesos Industriales” que está orientado al desarrollo general de sistemas de apoyo para la toma de decisiones en procesos industriales, está enfocado específicamente al estudio, optimización e implementación de controladores inteligentes aplicados a modelos de equipos y procesos industriales característicos de la región.

Se desarrolla en el ámbito de la Facultad de Ingeniería de la Universidad Nacional de Jujuy (UNJu) bajo la dirección del Mg. Ing. Sergio L. Martínez, en un entorno multidisciplinario, con el aporte de las cátedras de Inteligencia Artificial (carrera de Ingeniería Informática), Ingeniería de Procesos y Modelo y Simulación (carrera de Ingeniería Química). El proyecto se compone con los siguientes integrantes: Codirector Ing. Samuel Franco Domínguez; Asesor Dr. Enrique E. Tarifa; Docente Investigador Ing. Miguel A. Azar; Docente Investigador Ing. Jorge J. Gutierrez.

Está financiado por la Secretaría de Ciencia y Técnica y Estudios Regionales (SeCTER) de la Universidad Nacional de Jujuy (UNJu) y se identifica con el código interno D/0133.

Este proyecto forma parte del Programa de Investigación “Desarrollo de Sistemas de Soporte a la Toma de Decisiones” (dirigido por el Dr. Enrique Tarifa y codirigido por el Mg. Ing. Sergio L. Martínez) y está incorporado al Programa de Incentivos –Cód. 08/D134– dependiente de la Secretaría de Políticas Universitarias (SPU) del Ministerio de Educación de la Nación.

Introducción

El ámbito industrial emplea procesos manuales, semiautomáticos y automáticos. Casi todos estos procesos incorporan controladores o se asisten con subsistemas de control; en forma simplificada y según el tipo de controlador que se trate, cuentan con una variable de referencia (*set point*), un proceso de comparación y la toma de una decisión a través de una acción. En el caso de los controladores automáticos, aún frente a los grandes avances de la ingeniería de control, el uso de los clásicos PID se emplea en casi todos los ámbitos de la industria (Khodabakhshian y Edrisi, 2008).

Existe un amplio campo de desarrollo y aplicación que tienen los controladores convencionales que, usualmente basados en modelos, realizan un manejo estrictamente matemático de las variables de control. Frente a ello, los controladores basados en inteligencia artificial, con un amplísimo campo de investigación y desarrollo, tienden a marcar una actividad superadora frente a los primeros, principalmente por presentar estructuras más simples con la factibilidad de aplicar variables de control simbólicas, sencillas y naturales.

Estado del Arte

Entre los métodos más antiguos y relevantes para la sintonización de sistemas de control convencionales, tanto a lazo abierto y especialmente a lazo cerrado, se destacan los trabajos clásicos de Ziegler y Nichols (1942), Cohen y Coon (1953), Lopez *et al.* (1967), Kaya y Sheib (1988), Sung *et al.* (1996), entre otros de gran importancia.

La sintonización de sistemas de control inteligentes, y más específicamente aquellos basados en lógica fuzzy (Zadeh, 1965, 1978, 1994), se realiza mediante diversas configuraciones en las que se modifican parámetros intrínsecos tales como las funciones de pertenencia del sistema de control fuzzy y las ganancias intervinientes, entre otros.

Para este tipo de controladores, el proceso de sintonía se basa en la modificación de la forma y posición de las funciones internas de controlador FLC (*Fuzzy Logic Controller*). Para ello, la técnica PSO –optimización por enjambre de partículas– (Kenedy *et al.*, 1995) es un enfoque ideal como proceso de optimización.

Las alternativas de solución para una sintonía Fuzzy-PSO plantean esquemas estructurales que conviene citar. Uno de ellos es el abordaje que desarrolla Bouallègue *et al.* (2012), donde el diseño del controlador fuzzy es formulado como un problema de optimización con restricciones que es eficientemente resuelto por el algoritmo PSO.

Otras alternativas se reflejan en los trabajos de Welch *et al.* (2010), Permana *et al.* (2010), Talbi *et al.* (2011), Liu *et al.* (2007), Das y Ghosh (2013), Gharghory y Kamal (2013), Turanoglu *et al.* (2011) y Kumar y Sathavara (2015), en donde el algoritmo PSO modifica los parámetros de las funciones de pertenencia del controlador hasta encontrar el mínimo error

de estado estacionario (SSE). Una variante a este caso la describen Letting *et al.* (2010) y Rahmani *et al.* (2012), donde el objetivo es minimizar los principales parámetros del controlador tales como el tiempo de crecimiento (*rising time*), tiempo de ingreso y permanencia en la banda de error (*settling time*), sobrepico (*overshoot*) y error de estado estacionario (*steady state error*).

El planteo de modificar únicamente las ganancias de un controlador FLC es otra de las variantes que presenta Obaid *et al.* (2011) en su trabajo de diseño óptimo de un controlador FLC tipo PID. Este esquema también es el planteado por Ahmed *et al.* (2014), Hurel *et al.* (2012) y Ghoshal (2004).

El control a través de la optimización de la base de reglas del sistema fuzzy propuesto por Zhang *et al.* (2010) es otro caso en donde se aplican diferentes pesos a las reglas obteniendo resultados muy satisfactorios con respecto al sistema fuzzy sin pesos y a un sistema híbrido aplicando algoritmos genéticos (AG-Fuzzy). Bajo este esquema también Debnath *et al.* (2013) no solo actúa sobre la base de reglas sino también en forma simultánea aplica la optimización sobre las funciones de pertenencia.

Shayeghi *et al.* (2008; 2010) propone una estrategia en la que la optimización se aplica sobre los bloques de proporcionalidad, integral y diferencial.

La sintonización también es posible ajustando los parámetros de la función transferencia según expresa Al Gizi *et al.* (2012) en su trabajo *Improve Fuzzy-PSO PID Controller by Adjusting Transfer Function Parameters*, pero este enfoque es altamente cuestionable debido a que un sistema controlado (planta) posee características intrínsecas que no pueden ser alteradas en la práctica.

Líneas de Investigación, Desarrollo e Innovación

Del análisis de los diferentes trabajos y desarrollos relacionados con el proceso de sintonización de sistemas de control inteligentes, así como las diferentes pruebas realizadas aplicando el algoritmo PSO híbrido, se establece un procedimiento general de diseño de controladores fuzzy. Los pasos que se indican, según la profundidad y alcance con que sean considerados, pueden representar en sí mismos líneas complementarias de investigación:

- 1. Estudio de las estructuras de control:** esto incluye sistemas de lazo abierto/cerrado, respuestas ante impulso/escalón y modelado (Ogata, 2010; Jantzen, 2007).
- 2. Obtención de funciones de transferencia:** implica encontrar aquella función, dentro del campo de Laplace, que modele el comportamiento de la planta a controlar (Ogata, 2010; Love, 2007).
- 3. Determinación de las magnitudes de entrada/salida:** esto es, ponderar el alcance de las variables fuzzy que controlarán la planta de acuerdo a sus características operativas (Kavacic y Bogdan, 2006; Tanaka y Wang, 2001).
- 4. Diseño del algoritmo de optimización:** consiste en definir que variables serán candidatas a modificar para el ajuste para posteriormente codificar el sistema híbrido (Kumar y Sathavara, 2015; Gharghory y Kamal, 2013).
- 5. Pruebas de laboratorio y mediciones:** contrastar los resultados con casos reales factibles de ser medidos, como por ejemplo el control de motores de corriente continua o alterna.

En cuanto al desarrollo e innovación, el trabajo está orientado a modificar la técnica híbrida para un ajuste en tiempo real, dado que una planta (o sistema controlado)

puede resultar más compleja e incluso imposible de modelar con precisión. Es en estos casos que la sintonización requiere de un proceso dinámico y por lo tanto el modelado de la planta puede hacer uso de otras técnicas alternativas a las convencionales de forma analítica. Un caso de estas características se refleja en el trabajo de Azar *et al.* (2013), en donde se recurre a un modelado de planta bajo un esquema fuzzy.

Objetivos y Resultados

El proyecto de referencia propone como objetivo general *diseñar, sintonizar e implementar mediante simulación, sistemas de control inteligentes aplicados a modelos de procesos industriales enfocados hacia procesos productivos o tecnológicos de la región*, y se complementa con objetivos específicos que se detallan a continuación:

- Estudiar y modelar procesos industriales clásicos utilizados en procesos productivos.
- Desarrollar sistemas de control inteligentes basados en diversas tecnologías de inteligencia artificial.
- Aplicar procedimientos específicos para la generación de datos requeribles en el aprendizaje y configuración de sistemas de control inteligentes.
- Comparar y determinar las ventajas e inconvenientes que ofrecen los sistemas de control inteligentes frente a los sistemas clásicos de control.
- Investigar y aplicar nuevos métodos algorítmicos para mejorar la adaptabilidad de los sistemas de control inteligentes a los esquemas de procesos industriales modelados.

Respecto del trabajo que aquí se presenta, y de acuerdo a las primeras versiones aplicadas para un controlador basado en el algoritmo PSO-Fuzzy, se obtuvieron los siguientes resultados preliminares, utilizando como sistema controlado testi-

go, el modelo de un motor de corriente continua:

- El sistema de control, cuando es sometido a una perturbación de tipo escalón, responde a través de una curva amortiguada. Tal perturbación puede generarse, en el caso de un motor de corriente continua, mediante una carga agregada en forma brusca cuando el motor se encuentra en vacío y a velocidad nominal.
- La función objetivo a minimizar fue evaluada para las variables más sensibles del sistema de control al que se le aplica un escalón. Las variables seleccionadas fueron: tiempo de crecimiento (*rise time*), tiempo de ingreso y permanencia en la banda de error (*settling time*) y sobrepico (*overshoot*), para la función objetivo *FO* siguiente:

$$FO = RiseTime + SettlingTime + Overshoot$$

- El algoritmo PSO genera partículas aleatorias que se mueven dentro de un rango muy acotado. Cada partícula puede incursionar en el rango definido para cada partición de las funciones de pertenencia del controlador fuzzy.
- En este caso, las partículas del algoritmo PSO se mueven en un espacio multidimensional que es desconocido pero que es medible en forma indirecta a través de las variables mencionadas anteriormente.

Las primeras aplicaciones del algoritmo PSO-Fuzzy para sintonizar controladores FLC arrojaron resultados muy prometedores, pero el proceso de investigación continúa. Actualmente se trabaja con variantes de la estrategia ABC (algoritmo de colonia de abejas) para buscar las mejores propiedades en el proceso de sintonía y optimización de los sistemas, que luego serán aplicados a procesos industriales característicos de esta región.

Formación de Recursos Humanos

En los dos años de ejecución del proyecto se obtuvieron resultados que dieron lugar a trabajos presentados en congresos y publicaciones en revistas científicas. Además, se desarrollaron modelos matemáticos de procesos industriales, controlados con distintos tipos de sistemas, tanto convencionales como inteligentes.

Becas de Investigación

- Cristian David Yurquina, estudiante de Ingeniería Química de la UNJu, con beca de Estímulo a las Vocaciones Científicas, CIN, 2014-2015. Título del plan de trabajo “Control Inteligente para piletas de concentración de una planta procesadora de litio”. Director: Mg. Ing. Sergio L. Martínez, Codirector: Ing. Álvaro F. Nuñez.
- Lautaro Acosta, estudiante de Ingeniería Química de la UNJu, con beca de Estímulo a las Vocaciones Científicas, CIN, 2015-2016. Título del plan de trabajo “Modelado y control inteligente de un proceso de concentración de Litio mediante pozas solares”. Director: Mg. Ing. Sergio Luis Martínez, Codirector: Ing. Álvaro F. Nuñez.

Formación de Investigadores

- Ing. Miguel A. Azar, Facultad de Ingeniería, UNJu. Integrante del proyecto “Diseño y Simulación de Controladores Inteligentes Aplicados a Procesos Industriales” en carácter de Docente Investigador.
- Ing. Jorge J. Gutierrez, Facultad de Ingeniería, UNJu. Integrante del proyecto “Diseño y Simulación de Controladores Inteligentes Aplicados a Procesos Industriales” en carácter de Docente Investigador.

Referencias

- Ahmad A.H. and Sultan N.S., “Design and Implementation of Controlled Zeta Converter Power Supply”, American Journal of Electrical and Electronic Engineering, V.2(3), pp. 121-128. 2014.
- Al Gizi A.J.H. and Mustafa M.W., “Improve Fuzzy-PSO PID Controller by Adjusting Transfer Function Parameters”, International Journal of Scientific and Research Publications, Volume 2, Issue 11, ISSN 2250-3153. 2012.
- Azar M.A., Martínez S.L. y Manero J., “Modelado con lógica fuzzy de sistema controlado con función transferencia desconocida: un caso de estudio”, IX Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA, ISSN 1853-7871. 2013.
- Bouallègue S., Haggège J. and Benrejeb M., “A New Method for Tuning PID-Type Fuzzy Controllers Using Particle Swarm Optimization”, Fuzzy Controllers – Recent Advances in Theory and Applications. ISBN 978-953-51-0759-0. 2012.
- Cohen, G.H. and Coon, G.A., “Theoretical Considerations of retarded Control”, ASME Transactions (EUA), Vol. 75, pp. 827 – 834, 1953.
- Das, D. and Ghosh A., “Algorithm for a PSO Tuned Fuzzy Controller of a DC Motor”, International Journal of Computer Applications, Vol. 73, N° 4, 2013.
- Debnath S.B.C., Shill P.C. and Murase K., “Particle Swarm Optimization Based Adaptive Strategy for Tuning of Fuzzy Logic Controller”, International Journal of Artificial Intelligence & Applications (IJAIA), Vol.4, No.1, 2013.
- Gharghory S. and Kamal H., “Modified PSO for Optimal Tuning of Fuzzy PID Controller”, IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 1, 2013.
- Ghoshal S.P., “Optimizations of PID gains by particle swarm optimizations in fuzzy based automatic generation con-

- trol”, Electric Power Systems Research, Elsevier, 72 pp. 203–212. 2004.
- Hurel J., Mandow A. and Garcia-Cerezo A., “Tuning a fuzzy controller by particle swarm optimization for an active Suspension System”, ECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, pp. 25-28, 2012.
- Jantzen J., "Foundations of Fuzzy Control", John Wiley & Sons Ltd, ISBN: 0-470-02963-3, England, 2007.
- Kaya, A. y Sheib T.J., “Tuning of PID Controllers of Different Structures”, Control Engineering (EUA), pp. 62–65, 1988.
- Kennedy J. and Eberhart R., “Particle swarm optimization,” Proceedings of IEEE International Conference on Neural Networks, pp. 1942–1948. Vol. 4, 1995.
- Khodabakhshian and M. Edrisi, “A New Robust PID Load Frequency Controller”, Control Engineering Practice, Vol. 16, pp. 1069-1080, 2008.
- Kovacic Z., Bogdan S., "Fuzzy Controller Design, Theory and Applications", ISBN 978-0-8493-3747-5, CRC Press, Boca Raton, 2006.
- Kumar M.H. and Sathavara M.J., “Comparisons of Speed Control of DC Motor Using PID-PSO and Fuzzy-PSO Techniques”, International Journal For Technological Research In Engineering Vol. 2, Issue 9, ISSN (Online): 2347–4718, 2015.
- Letting L.K., Munda J.L. and Hamam A., “Particle Swarm Optimized T-S Fuzzy Logic Controller for Maximum Power Point Tracking in a Photovoltaic System”, in IPEC, 2010 Conference Proceedings, pp. 89–94, 2010.
- Liu H., Abraham A. and Zhang W., “A fuzzy adaptive turbulent particle swarm optimization”, Int. J. Innovative Computing and Applications, Vol. 1, No. 1. 2007.
- López, A.M., Miller J.A., Smith C.L. and Murril P.W., “Tuning Controllers with Error-Integral Criteria”, Instrumentation Technology (EUA), 1967.
- Love J., "Process Automation Handbook", Springer-Verlag London Limited, ISBN 978-1-84628-281-2, England, 2007.
- Obaid Z.A., Salman S.A.A., Ali H.I., Sulaiman N., Marhaban M.H. and Hamidon M.N., “Design of PSO-Based Optimal Tunable PID Fuzzy Logic Controller Using FPGA”, A Ubiquitous Tool for the Practical Engineer, Prof. Clara Ionescu (Ed.), ISBN: 978-953-307-907-3, InTech, 2011.
- Ogata K., “Ingeniería de Control Moderna”, Prentice Hall, Hispanoamericana S.A. Minnesota, USA. 2010.
- Permana K.E. and Hashim S.Z.M., “Fuzzy Membership Function Generation using Particle Swarm Optimization”, Int. J. Open Problems Compt. Math., Vol. 3, No. 1, ISSN 1998-6262, 2010.
- Rahmani R., Mahmodian M.S., Mekhilef S. and Shojaei A.A., "Fuzzy logic controller optimized by particle swarm optimization for DC motor speed control". IEEE Student Conference on Research and Development. pp. 109-113. 2012.
- Shayeghi H., Jalili A. and Shayanfar H.A., “Multi-stage fuzzy load frequency control using PSO”, Energy Conversion and Management 49 pp. 2570–2580. 2008.
- Shayeghi H. and Shayanfar H.A., “PSO Based Neuro-Fuzzy Controller for LFC Design Including Communication Time Delays”, International Journal on “Technical and Physical Problems of Engineering” (IJTPE), Iss. 3, Vol. 2, No. 2, Jun. 2010.
- Sung, S.W., O J., Lee I.B., Lee J. and Yy S.H., “Automatic Tuning of PID Controller using Second-Order plus Time delay Model”, Journal of Chemical Engineering of Japan, Vol. 29 N° 6, pp. 990– 999, Japan, 1996.
- Talbi N. and Belarbi K., “Fast Hybrid PSO and Tabu Search Approach for Optimization of a Fuzzy Controller”, IJCSI International Journal of Computer Sci-

- ence Issues, Vol. 8, Issue 5, No 2, ISSN (Online): 1694-0814, 2011.
- Tanaka K., Wang H., "Fuzzy control systems, design and analysis", John Wiley & Sons, Inc., New York, 2001.
- Turanoglu E., Ozceylan E. and Kiran M.S., "Particle Swarm Optimization and Artificial Bee Colony Approaches to Optimize of Single Input-Output Fuzzy Membership Functions", Proceedings of the 41st International Conference on Computers & Industrial Engineering, pp. 542-547, 2011.
- Welch R.L. and Venayagamoorthy G.K., "Energy dispatch fuzzy controller for a grid-independent photovoltaic system", Energy Convers Manage, 51:928–937. 2010.
- Zadeh L., "Fuzzy sets as a basis for theory of possibility", Fuzzy Sets and Systems, vol. 1, pp. 3–28, 1978.
- Zadeh L., "Fuzzy sets," Information Control, vol. 8, pp. 338–353, 1965.
- Zadeh, L., "Soft computing and fuzzy logic", IEEE Software, 11(6), pp. 48-56. 1994.
- Zhang L., Gao L., Shao X., Wen L. and Zhi J., "A PSO-Fuzzy group decision-making support system in vehicle performance evaluation", doi:10.1016 / j.mcm.2010.03.042. 2010.
- Ziegler, J.B. and N.B. Nichols; Optimum Settings for Automatic Controls, ASME Transactions (EUA), Vol. 64, pp. 759-768, 1942.

Robótica y Fenotipado de Alta Capacidad con Relevamiento de Datos en Campo. Aplicaciones en Agricultura de Precisión.

Eduardo Álvarez¹, Sandra Serafino¹, Benjamin Cicerchia², Claudia Russo¹,
Mónica Sarobe¹, Pablo Luengo¹, Adrian Jaszczyszyn¹, Hugo Ramón¹

¹Instituto de Investigación y Transferencia en Tecnología (ITT), Escuela de
Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires
(UNNOBA)

²Becario de la Comisión de Investigaciones Científicas de la Pcia. De Bs. As.
(CIC)

Sarmiento y Newbery (CP 6000), Junín, Buenos Aires, Argentina. Teléfonos (0236) 4636945/44

{eduardo.alvarez, sandra.serafino, benjamín.cicechia, claudia.russo, monica.sarobe, pabloluengom
adrianjaz, hugo.ramon}@itt.unnoba.edu.ar

Resumen

En la UNNOBA se está trabajando en la aplicación combinada de diferentes desarrollos tecnológicos y Agricultura de Precisión. Ambas áreas de estudio constituyen una herramienta fundamental para lograr un manejo adecuado y preciso del suelo y sus cultivos en base a su variabilidad dentro de un lote, permitiendo adaptarse a las exigencias de la agricultura moderna en el manejo óptimo de grandes extensiones.

El uso de herramientas tecnológicas orientadas al uso de imágenes y sensores (GPS, sensores, UAVs) en la agricultura de precisión, permite diferenciar variabilidad y características particulares de diferentes coberturas terrestres para mejorar la toma de decisiones en pos de obtener mayores rendimientos. En la actualidad se encuentran muchos trabajos de investigación que utilizan imágenes de sensado remoto (satélites, áreas)[1][2][3].

La presente línea de investigación pretende aportar desde otra perspectiva mediante el uso plataformas robóticas de sensado a campo y el uso de imágenes digitales capturadas con cámaras de luz visible, multi o hiper espectrales, térmicas, más la utilización de técnicas de procesamiento digital, con el fin de aportar un valor agregado a las tecnologías ya existentes en esta área. Esto permitirá mejorar el estudio de aspectos cualitativos y cuantitativos de diferentes tipos de cultivos en relación a su variabilidad fenológica, morfológica, fisiológica, temporal y espacial.

Palabras clave: Agricultura, Fenotipado, Robótica, Imágenes.

Contexto

Esta línea de investigación forma parte del proyecto “Tecnologías exponenciales en contextos de realidades mixtas e interfaces avanzadas.” aprobado por la

Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA en el marco de la convocatoria a Subsidios de Investigación Bianaules (SIB2015). A su vez se enmarca en el contexto de planes de trabajo aprobados por la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires y por la Secretaría de Investigación de la UNNOBA en el marco de la convocatoria “Becas de Estudio Cofinanciadas 2015 CIC Universidades del interior bonaerense”.

A esto se suma el proyecto “Fenotipado de Alta Capacidad con Relevamiento de Datos en Campo” el que se encuentra actualmente en desarrollo. Este es un trabajo conjunto entre la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA y el INTA, Estación Experimental Pergamino.

El proyecto se desarrolla en el Instituto de Investigación en Tecnologías y Transferencia (ITT) dependiente de la mencionada Secretaría, y se trabaja en conjunto con la Escuela de Tecnología de la UNNOBA.

El equipo está constituido por profesionales de ambas instituciones, UNNOBA e INTA. Por parte de la UNNOBA intervienen docentes e investigadores pertenecientes al ITT, al Departamento de Ciencias Agrarias, Escuela de Ciencias Agrarias así como también, estudiantes de las carreras de Informática de la Escuela de Tecnología de la UNNOBA. Y por parte del INTA, intervienen; Mejoradores Genéticos, Ecofisiólogos e Ingenieros Agrónomos.

Introducción

La Agricultura de Precisión es el resultado del avance de la tecnología en todo lo que sea siembra de precisión, relevamiento de datos a campo y manejo de grandes volúmenes de datos en un tiempo limitado, tal que las conclusiones obtenidas sean aplicables al siguiente ciclo de siembra. La condición de relevamiento a campo es excluyente, ya que está demostrado que los desvíos que se producen en mediciones hechas en cultivos en ambientes controlados hacen inviables su extrapolación a situaciones de siembra en condiciones reales de producción. Aquí, las aplicaciones basadas en robótica móvil tanto aérea como terrestre, toma y procesamiento de imágenes, y procesamiento masivo de datos, juegan un papel central. [3,4]. Necesitamos entonces desarrollar Plataformas de Fenotipado de alto rendimiento y gran caudal de datos.[1]

En relación al desarrollo tecnológico en esta área, tanto el sensado remoto de imágenes, como el sensado de otro tipo de señales (temperatura, distancia, posicionamiento, etc.) han sido y siguen siendo un aporte muy importante [5,6].

En particular, en el área de imágenes, los mayores desafíos, entre otros, se encuentran en el procesado y análisis de grandes volúmenes de datos compuestos obtenidos de diferentes sensores, y mejorar la resolución espacial lo que nos va a permitir trabajar con mayor nivel de detalle (necesario para el tipo de datos que se desean relevar). Dado que el tipo de procesamiento de imagen requerido, está fuertemente ligado a la naturaleza del rasgo buscado, la resolución con la que pueda captarse la escena y sus objetos de interés es muy importante. Esto impacta directamente en el correcto estudio y

selección de los sensores y mecanismos utilizados para su recolección [7]. Mejorar los mecanismos de captura de datos redundará sin duda en la reducción de los costos computacionales asociados al preprocesamiento y al procesamiento y posterior análisis, propiamente dicho.

Particularmente optaremos por no utilizar sensado remoto, sino desarrollar una plataforma robótica que nos permita garantizar esta condición, con el objetivo de mejorar el procesamiento y análisis posterior de la escena obteniendo los resultados deseados a campo.

La información a recolectar incluye datos sobre el rendimiento del cultivo, nutrientes y otras propiedades del suelo, nutrientes de los cultivos, volumen del follaje y la biomasa, contenido de agua, y posible existencia de plagas (enfermedades, malezas e insectos), entre otros.

Una vez obtenido un conjunto de imágenes normalizadas (misma escena, con variabilidad temporal, resolución espacial requerida, diferentes estados evolutivos del cultivo, etc.), se aplicarán en función del resultado esperado, diferentes técnicas de procesamiento digital, sobre la base de conocimiento aportada por los expertos. Estas técnicas se seleccionarán teniendo en cuenta tanto la problemática particular a resolver, como el volumen de datos recolectado, optimizando los costos computacionales involucrados. No sólo se desarrollarán las herramientas tecnológicas necesarias sino que se utilizarán todas aquellas existentes que aporten a la solución requerida como por ejemplo, el uso de diferentes tipos de índices de vegetación, (NDVI-Normalized Difference Vegetation Index-, SAVI-Soil Adjusted Vegetation Index -, entre otros) [8].

En lo referente al uso de plataformas móviles para fenotipado a campo se debe tener en cuenta que la plataforma debe tener capacidades tales como medición de variables ambientales como ser temperatura y humedad, sensado de presencia de vegetación para el caso de utilizar guiado por imagen, capacidad de ubicación y orientación espacial, autonomía de marcha garantizada con el uso de paneles solares, sistema de guía y captura de imagen posible de operar en modo de trabajo manual, semiautomático y automático.

Dado el volumen de información esperado, se prevé que la plataforma tendrá capacidad de captura, procesamiento y almacenamiento de información en soportes removibles. La plataforma además debe ser de operación segura para la vida humana y todos sus dispositivos operar con baja tensión de alimentación. Otras condiciones a obtener para todos los sistemas de la plataforma son; la facilidad de operación, la replicabilidad constructiva, bajo costo de operación y mantenimiento, y en general, confiabilidad para todos los mecanismos de hardware y software involucrados. No se descarta el hecho de necesitar operar varias plataformas en forma simultánea para acortar los intervalos de medición pudiendo cubrir más de un ensayo como alternativa a distintas propuestas comerciales disponibles hoy en el mercado de muy alto costo [10]. En el caso de requerir tomas aéreas complementarias, se prevee además el uso de naves no tripuladas como multicópteros y alas volantes.

Líneas de Investigación, Desarrollo e Innovación

Las líneas de investigación que se proponen son:

- Diseño y fabricación de plataformas de fenotipado a campo.
- Navegación autónoma en ambientes no controlados (campos).
- Diseño y desarrollo de redes de sensores de bajo consumo para captura de datos.
- Desarrollo y normalización de un proceso de captura de datos de diferentes sensores.
- Estudio y desarrollo de técnicas de procesamiento y análisis digital de imágenes para grandes volúmenes integrados de datos.

Se proyectan además trabajos futuros sobre estas líneas de investigación, en el marco del Programa Iberoamericano CYTED (Ciencia y Tecnología para el Desarrollo) [9] y la Red Iberoamericana para fortalecer la sostenibilidad agroalimentaria mediante el desarrollo de la Inteligencia Organizacional (RISADIO), en conjunto con Cuba.

Resultados y Objetivos

El avance de la tecnología en los últimos años ha ido abarcando diferentes disciplinas, entre las cuales la agricultura no es la excepción. En la actualidad la Agricultura de Precisión ha ido incorporando tecnologías de la información y la comunicación para dar soporte a diferentes mediciones.

El estudio fenotípico como parte de la AP ha comenzado a utilizar la tecnología para la medición de las diferentes características que el cultivo posee lo que

lo hace un campo de investigación con un gran potencial de desarrollo a futuro. En este sentido surgen constantemente nuevas mediciones y la necesidad de la utilización de diferentes tecnologías para resolver nuevos desafíos.

Como resultado del trabajo de investigación y desarrollo indicado en las líneas mencionadas, se espera obtener una plataforma de *fenotipado de alto caudal a campo* que sea sencilla, fácil de operar, que sea replicable y que tenga el menor costo posible, y la mayor disponibilidad de sensores [11].

Formación de Recursos Humanos

En esta línea de I/D se espera concluir con un trabajo de Prácticas Profesionales Supervisadas de la Ingeniería en Informática e iniciar la dirección, realización de dos trabajos de posgrado, y una Beca de Estudio Cofinanciada otorgada por la **Comisión de Investigaciones Científicas (CIC) y la UNNOBA**. Asimismo se espera desarrollar dos tesis doctorales y dos tesinas de grado, dirigidas por miembros de este proyecto

Referencias

- [1] Araus J., Cairns, Jill. (2014). **Field high-throughput Phenotyping: the new crop breeding frontier**. Department of plant Biology, Unit of Plant Physiology, University of Barcelona, Spain. CIMMYT Southern Africa Regional Office, Harare, Zimbabwe.
- [2] Breccia G., Nestares G. **Next-generation phenotyping in plants: old problems, new promises**. Facultad de Ciencias Agrarias, Universidad Nacional de Rosario. Consejo Nacional de

- Investigaciones Científicas y Técnicas (CONICET).*
- [3] Delrieux C., et al. (2014). **Ortomosaicos utilizando Imágenes Aéreas tomadas por Drones y su aplicación en la Agricultura de Precisión.** Laboratorio de Ciencias de las Imágenes, Universidad Nacional del Sur. EEA Manfredi.
- [4] Delalieux S. et al. (2014). **Unmixing-Based Fusion of Hyperspatial and Hyperspectral Airborne Imagery for Early Detection of Vegetation Stress.** *IEEE Journal of selected topics in applied earth observations and remote sensing*, VOL. 7, NO. 6, JUNE 2014
- [5] Berni J.A.J. et al. **Remote sensing of vegetation from UAV platforms using light weight multispectral and thermal imaging sensors.** *Quantalab, Instituto de Agricultura Sostenible (IAS), Consejo Superior de Investigaciones Científicas (CSIC), 14004 Cordoba, Spain*
- [6] Calderón R. et al. (2014). **Detection of downy mildew of opium poppy using high-resolution multi-spectral and thermal imagery acquired with an unmanned aerial vehicle.** *Precision Agriculture. An International Journal on Advances in Precision Agriculture. ISSN 1385-2256. Volume 15. Number 6.*
- [7] Lee, W.S. et al. (2010) **Sensing technologies for precision specialty crop production.** *Computers and Electronics in Agriculture.* 74, 2-33.
- [8] Díaz García-Cervigón, J. J. (2015) **Estudio de Índices de vegetación a partir de imágenes aéreas tomadas desde UAS/RPAS y aplicaciones de estos a la agricultura de precisión.** *Universidad Complutense de Madrid*
- [9] CYTED. <http://www.cyted.org>
- [10] LEMNATEC. <http://www.lemnatec.com>
- [11] MSc. Torres Clayton J, et al (2012). **Kinematic study of an agricultural robotic platform.** *Universidade de São Paulo. Brasil. Universidad Industrial de Santander. Colombia.*

Planificación de tareas en Sistemas de Tiempo-Real

Claudio Aciti^{1,2}, Martín Rodríguez, Guillermo Arispe, Martín Beltrame, Mariano Castellano, Martín Errobidarth, Martín Fernández Gamen, Diego Gariboldi, Ignacio Luna, Santiago Rojo y Diego Tabares

¹ Departamento de Ciencia y Tecnología
Universidad Nacional de Tres de Febrero (UNTREF)
Valentín Gómez 4752 (+54-011-47593528) - Caseros - Buenos Aires

² Departamento de Computación y Sistemas
Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA)
Pinto 399 (+54-249-4385680) - Tandil - Buenos Aires
caciti@exa.unicen.edu.ar

Resumen

La presente línea de investigación promueve la temática de Sistemas de Tiempo Real en la carrera Ingeniería en Computación de la UNTREF. Preferentemente se profundiza el tema planificación de tareas de tiempo-real, tanto en mono/multi procesadores, como procesadores heterogéneos. Al ser una carrera sin egresados, se cuenta solo con alumnos, los cuales ya están abordando la implementación de una herramienta de simulación de tareas de sistemas de tiempo real con procesadores heterogéneos y la implementación de sistemas reales en diferentes plataformas.

Se pretende impulsar el desarrollo de las industrias, ubicadas en la región, vinculadas a proyectos de tipo Hardware/Software y fortalecer vínculos interuniversitarios.

Palabras clave: sistemas de tiempo real - sistemas embebidos - planificación de tareas

Contexto

La línea “Planificación de tareas de tiempo-real” está enmarcada dentro del proyecto interno de investigación “Técnicas de Aceleración en Arquitecturas Heterogéneas de Computación” radicado en UNTREF. Dicho proyecto es nuevo y forma parte de la reestructuración de la carrera Ingeniería en Computación. Participan tres profesores con posgrados y alumnos de la carrera. Además, se mantiene un vínculo directo con docentes y alumnos de la Universidad Nacional del Centro de la Provincia de Buenos Aires. El financiamiento está íntegramente a cargo de la UNTREF.

Introducción

Las nuevas tendencias del mercado junto con la baja capacidad de innovación tecnológica que poseen las PyMEs a nivel regional, plantean una problemática crucial para su sustentabilidad en el

mediano plazo. Uno de los pocos caminos posibles que las PyMEs tienen de acceder a procesos de innovación y transformación de la actividad, es asociándose entre sí y generar proyectos conjuntos con otras industrias e instituciones vinculadas al sistema científico tecnológico. Para ello, es necesario generar un conjunto de capacidades a las que concurran las habilidades, conocimientos, experiencias y talentos multidisciplinarios existentes en el país, para encontrarse y colaborar efectivamente en la generación de tecnologías, productos, servicios y procesos competitivos con proyección de largo plazo. Este proyecto pretende impulsar el desarrollo de sistemas de tiempo-real en las industrias ubicadas en la región.

Un sistema de tiempo-real (STR) es aquel que no solo depende de la corrección del resultado sino también del cumplimiento de un plazo de tiempo preestablecido. En general, los STR interactúan con el entorno que los rodea. En consecuencia, el cumplimiento de los plazos de tiempo es condición excluyente para su correcto funcionamiento [1]. Los STR se aplican en áreas tales como control digital, procesamiento de señales, sistemas de telecomunicación y sistemas multimedia, entre otros. Cada vez más, los sistemas informáticos utilizan STR empotrados, que cumplen con una tarea determinada de un sistema mayor. Esto conlleva una limitación de recursos (potencia de procesador, memoria, interfaces gráficas), y la necesidad de ejecutar tareas de tiempo-real (con distintas prioridades) concurrentemente [2][3][4]. La ejecución de estas tareas, en conjunto, hacen que la construcción de un STR tenga características específicas y diferentes, respecto a otros sistemas informáticos [5][6][7][8][9]. En los últimos años está teniendo un auge muy

importante debido a los cambios tecnológicos ocurridos, que permiten incorporar sistemas empotrados de tiempo-real en casi cualquier dispositivo electrónico, como parte de un sistema informático más grande. La necesidad de dotar a estos dispositivos de inteligencia y de que cumplan plazos de tiempo en sus respuestas (navegadores con GPS; teléfonos móviles, vehículos navegadores terrestres, aéreos y submarinos; robots bípedos, cuadrúpedos; etc.) [10][11] hace necesario que los futuros egresados de la UNTREF tengan una base de conocimiento que les permita desenvolverse de forma adecuada en el desarrollo de este tipo de sistemas.

Líneas de Investigación, Desarrollo e Innovación

Dicha línea de investigación recién inicia dentro de la universidad, por lo tanto es básica y los desarrollos son experimentales en el área de Sistemas de Tiempo Real. Está particularmente enfocado a la planificación de tareas con mono/multi procesadores y procesadores heterogéneos. Se está armando un laboratorio experimental que, al momento, cuenta con placas del tipo edu-CIAA, arduino, raspberry Pi, Intel Galileo y kits de freescale. El principal campo de aplicación, a mediano plazo, es la producción y tecnología de dispositivos de propósito dedicado y aquellos de propósito general con requerimientos de tiempo real que se comercialicen en industrias de la región.

Resultados y Objetivos

Si bien al momento no hay resultados directos, se está trabajando en el desarrollo de un prototipo de un

simulador de tareas de tiempo real con mono/multi procesadores y procesadores heterogéneos. Se espera en el transcurso de 2016 tener el prototipo funcionando. También se están realizando prácticas con la placa edu-CIAA, y se espera durante este año tener varios desarrollos experimentales en marcha. Además, se prevé el dictado de cursos introductorios para profesionales de la región con las plataformas Intel Galileo y edu-CIAA.

Formación de Recursos Humanos

El equipo de Sistemas de Tiempo Real es dirigido por un profesor con posgrado, quien se especializa en el tema “Análisis frecuencial de Sistemas de Tiempo Real en aplicaciones de control industrial”. También lo integran 3 alumnos de la UNTREF, uno de ellos lleva adelante una Beca de Entrenamiento (CIC) durante 2016 bajo el título “Simulador de planificador de tareas de tiempo-real con procesadores heterogéneos” y será su tesina de grado. También participan 5 alumnos quienes desarrollan el trabajo final de la materia Sistemas de Tiempo-Real y están relacionados al proyecto. Se espera que al menos dos continúen con el tema y presenten su tesina de grado.

Por otra parte, participan 2 alumnos quienes desarrollan sus trabajos finales de grado en la UNICEN y están directamente ligadas a esta línea y se prevé su exposición en el transcurso de este año.

Referencias

[1] P. A. Laplante, *Real-Time Systems Design & Analysis*. Wiley India Pvt. Ltd. (2006)

[2] *Sistemas de Tiempo Real y Lenguajes de Programación*, Burns A., Wellings A., 3ra

Edición, Addison Wesley, ISBN 84-7829-058-3. (2003)

[3] *Real-Time Systems*, Liu, J.W.S., Prentice Hall. ISBN 0-13-099651-3. (2000)

[4] *Real-Time Computer Control. An Introduction*, Bennet S., 2da Edición, Prentice Hall, ISBN 0-13-764176-1. (1994)

[5] G. Buttazzo. Rate Monotonic vs. EDF: Judgment Day. *Real-Time Systems*, Vol. 29, Issue 1, pp. 5-26. (2005)

[6] Sha, L., Abdelzaher, T., Árzn, K-E., Cervin, A., Baker, T., Burns, A., Caccamo, M., Lehoczky, J. and Mok, A. K. “Real Time Scheduling Theory: A Historical Perspective”. *Real-Time Systems* 28(2-3):101-155. (2004)

[7] G. Bernat, and R. Cayssials. “Guaranteed on-line weakly-hard realtime systems”. In *proc. 22nd IEEE Real-Time Systems Symposium*, pages 25-35, London. (2001)

[8] G. Buttazzo, M. Bertogna, and G. Yao. “Limited Preemptive Scheduling for Real-Time Systems. A Survey”. *IEEE Transactions on Industrial Informatics*, Vol. 9, No. 1, (2012)

[9] L. Sha, R. Rajkumar, and J. Lehoczky, “Priority inheritance protocols. An approach to real-time synchronization”. *IEEE Trans. Comput.*, vol. 39, no. 9, pp. 1175–1185. (1990)

[10] T. P. Baker, “Stack-based scheduling for realtime processes,” *Real-Time Syst.*, vol. 3, no. 1, pp. 67–99. (1991)

[11] E. Bini and G. C. Buttazzo, “Schedulability analysis of periodic fixed priority systems,” *IEEE Trans. Comput.*, vol. 53, no. 11, pp. 1462–1473. (2004)

Un Modelo para la Evaluación de la Seguridad en Sistemas Informáticos

Aristides Dasso, Ana Funes

SEG / Departamento de Informática / Facultad de Ciencias Físico-Matemáticas y Naturales / Universidad Nacional de San Luis
Ejército de los Andes 950, D5700HHW San Luis, Argentina
+54 (0) 266 4520300, ext. 2126
{arisdas, afunes}@unsl.edu.ar

Resumen

Dentro del contexto de desarrollo de modelos de evaluación de sistemas complejos, esta investigación tiene como objetivo el concretar un modelo que permita evaluar el nivel de seguridad de sistemas informáticos. Para ello, comenzamos por establecer un conjunto de características (los requisitos de seguridad) en formato jerárquico, tomado de la norma ISO 2700, para luego, aplicando el método de evaluación Logic Scoring of Preference (LSP), construir un modelo adecuado que permita obtener un resultado numérico final entre 0 y 100 el cual indique claramente cuál es el grado de seguridad del sistema bajo evaluación.

Palabras clave: Seguridad de Sistemas Informáticos. Evaluación de la Seguridad de Sistemas Informáticos. Métodos de Evaluación. Logic Scoring of Preference (LSP).

Contexto

Este trabajo de investigación se viene llevando a cabo dentro del SEG (Software Engineering Group), en el ámbito de la Universidad Nacional de San Luis y se

encuentra enmarcado dentro de una de las líneas de investigación del Proyecto de Incentivos código 22/F222 “Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software” (Director: Daniel Riesco, Co-Director: Roberto Uzal. Acreditado con evaluación externa. Financiamiento: Universidad Nacional de San Luis).

Introducción

La construcción de modelos de evaluación de sistemas complejos, entre los que se encuentran los sistemas de seguridad informática, constituye una necesidad primordial para garantizar que las medidas, herramientas, métodos, etc., tomadas en ese sentido sean las más adecuadas en un entorno dado.

Existen propuestas, no sólo de cómo organizar e implementar la seguridad de sistemas informáticos, sino también de cómo se puede certificar la seguridad de los mismos. Por ejemplo Common Criteria Schemes [13][14] es un método que armoniza criterios sobre seguridad de productos de software y hardware que se aplica en varios países. Ha recibido algunas críticas ya que las certificaciones que otorga pueden ser parciales (ver por

ejemplo [15], [16]).

Se encuentran, además, en la literatura, en empresas y en organismos gubernamentales, numerosas referencias a modelos, métricas, y sistemas de evaluación de la seguridad (ver por ejemplo [17], [18], [19], [20], [21]).

Cabe aclarar que el método LSP (Logic Score of Preferences) [1], [6], [7], [8], que adoptamos para construir nuestros modelos de evaluación, es un método que se basa en el empleo de una lógica continua, que permite la creación de funciones complejas de evaluación y su aplicación en la evaluación y comparación de sistemas de índole general, permitiendo la creación de modelos precisos y fácilmente adaptables a las necesidades del usuario, en este caso las necesidades de Seguridad de un sistema informático.

El proceso general propuesto por el método LSP es mostrado en la Figura 1. El desarrollo e identificación de la lista de características principales a tener en cuenta (requisitos) corresponde al primer nivel del *árbol de requerimientos* que el método prescribe construir en una de sus etapas.

Cada una de estas características del primer nivel comprende varios ítems o categorías más específicas de acuerdo con propiedades similares, tales que las mismas puedan ser razonablemente agrupadas, y así sucesivamente, llegando hasta ítems que no se desagregan, donde encontramos las ‘hojas’ del árbol de requerimientos, llamadas *variables de performance*.

Las hojas del árbol de preferencias son empleadas para construir una *estructura de agregación* junto con los operadores de Lógica Continua provistos por el método. Estos operadores o funciones GCD (Generalized Conjunction Disjunction) nos permiten agregar los valores observados de cada una de las

variables de performance, previamente mapeados a valores en el intervalo [0, 100] llamados *preferencias elementales*, por medio de funciones llamadas *criterios elementales*. Las preferencias elementales representan el grado de cumplimiento con un requisito del sistema bajo evaluación.

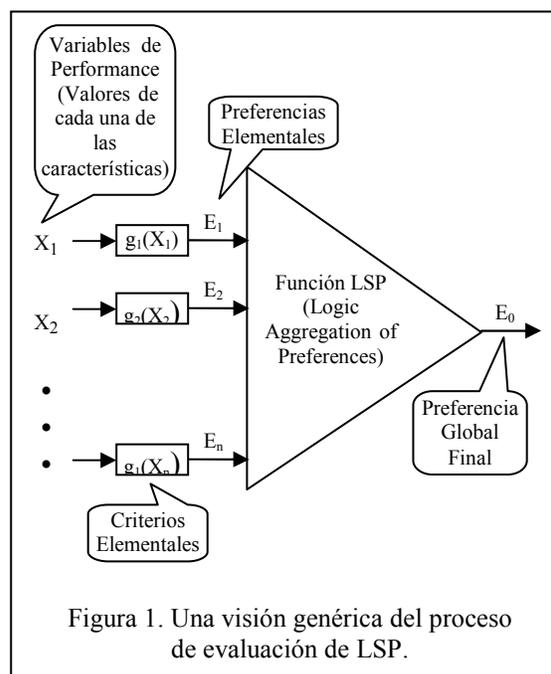


Figura 1. Una visión genérica del proceso de evaluación de LSP.

La estructura de agregación tiene como objetivo obtener, como resultado final un único valor (preferencia global final E_0 en la Figura 1) entre 0 y 100 que representa el grado de adecuación o de satisfacción de todas las características consideradas para el sistema bajo evaluación.

Así, por ejemplo, para alguna de las características que un sistema debería poseer, el valor asignado a la correspondiente variable de performance corresponderá a la valoración que se haga del mismo; dicho valor se transformará, con el correspondiente criterio elemental, en un valor del intervalo [0,100]. La relación entre el valor asignado a la variable y el intervalo será justamente propia de la elección del criterio elemental por parte de quienes construyan

el modelo.

Líneas de Investigación, Desarrollo e Innovación

La línea de investigación en la que se enmarca el trabajo presentado, es parte de una línea de investigación sobre el tema de la construcción de modelos de evaluación de sistemas complejos y que viene desarrollándose desde hace tiempo en el ámbito del SEG (Software Engineering Group), donde se han obtenido resultados plasmados en diversas publicaciones (ver por ejemplo [2], [3], [4], [10], [11], [12]).

Resultados y Objetivos

En una primera etapa nos encontramos desarrollando un modelo de evaluación que sigue de manera general las directivas establecidas en la norma ISO/IEC 27002. "Information technology - Security techniques - Code of practice for information security controls" [5]. Esta norma establece, como el título sugiere, un código de prácticas. El estándar tiene catorce cláusulas que en total establecen treinta y cinco categorías principales y ciento catorce controles.

Tanto las cláusulas, como las categorías y los controles tienen definidos sus objetivos y condiciones de ejecución e implementación, además de otra información útil para su ejecución.

En la Figura 2 mostramos, a modo de ejemplo, un árbol de preferencias parcial construido a partir de una de las catorce cláusulas, en este caso la cláusula "5. Access control", la que ha sido desagregada en sus cuatro categorías. Hay que remarcar que cada una tiene sus correspondientes controles (que no se muestran aquí por razones de espacio).

- 5. Access control.
 - 5.1. Business requirements of access control.
 - 5.2. User access management.
 - 5.2.1. User registration and de-registration.
 - 5.2.2. User access provisioning.
 - 5.2.3. Management of privileged access rights.
 - 5.2.4. Management of secret authentication information of users.
 - 5.2.5. Review of user access rights.
 - 5.2.6. Removal or adjustment of access rights.
 - 5.3. User responsibilities.
 - 5.4. System and application access control.

Figura 2. Variables de performance para la cláusula 5 [5].

Notemos que el árbol de preferencias y las variables de performance consideradas es una elección de quién o quienes construyen el modelo sobre la base de las necesidades del usuario.

A partir de un árbol de requerimientos, se pueden generar diversas estructuras de agregación como modelos de evaluación, previa clasificación de los distintos aspectos que el usuario considere mandatorios, opcionales y deseables para un sistema de seguridad. En la Figura 3 mostramos la estructura de agregación para el ítem "5.2. User access management."

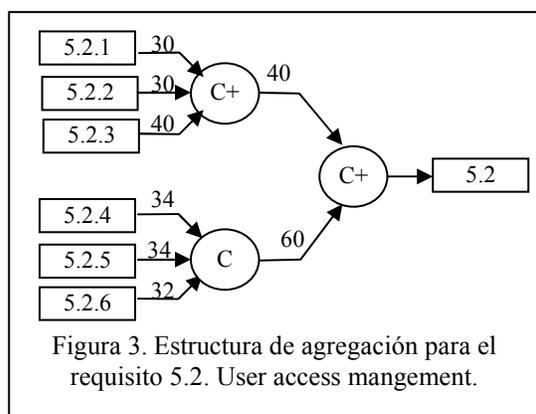


Figura 3. Estructura de agregación para el requisito 5.2. User access mangement.

Para asignar un valor a cada una de las variables de performance se sigue lo

prescripto por la norma en cada ítem (ver para este caso [5], págs. 21 y siguientes).

Como parte del trabajo futuro, esperamos, en una etapa siguiente, calibrar los modelos producidos, además de aplicarlos a casos reales. Asimismo, creemos que el modelo debería incluir la evaluación de los costes económicos de los aspectos de seguridad de los sistemas. Establecer el coste/beneficio de la seguridad es un aspecto sumamente importante para los usuarios.

También nos encontramos trabajando en la generación de un cuestionario para las empresas proveedores de sistemas de seguridad informática en todos sus niveles, con el objeto de obtener mayor información sobre las características de los sistemas ofrecidos, con el objeto de ampliar y/o mejorar nuestros modelos.

Formación de Recursos Humanos

Dentro del SEG (Software Engineering Group), en el ámbito de la Universidad Nacional de San Luis, en el que se realiza el Proyecto de Incentivos código 22/F222 “Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software”, se han llevado a cabo numerosas tesis de grado y de posgrado.

Entre otros, nos hemos concentrado en la evaluación de sitios de gobierno electrónico lo que ha dado como resultado una tesis de maestría en 2010; mientras que hay otras dos en preparación. La construcción del modelo aquí expuesto, también, tiene como objetivo ser motivo de tesis, como lo han sido la construcción de otras herramientas en el ámbito del proyecto.

Referencias

[1] Jozo J. Dujmovic, “Continuous Preference

Logic for System Evaluation”, IEEE Transactions on Fuzzy Systems, Vol. 15, N° 6, December 2007

- [2] A. Dasso, A. Funes, M. Peralta, C. Salgado, “Una Herramienta para la Evaluación de Sistemas”, Workshop de Investigadores en Ciencias de la Computación, WICC 2001, Universidad Nacional de San Luis, San Luis, Argentina, May 2001.
- [3] Ana Funes, Aristides Dasso, “Web Application Frameworks Evaluation”, CONAISI 2014, 13 y 14 de noviembre de 2014, San Luis, Argentina. pp. 1063-1070. ISSN: 2346-9927.
- [4] Ana Funes, Aristides Dasso, Carlos Salgado, Mario Peralta, “UML Tool Evaluation Requirements”. Argentine Symposium on Information Systems ASIS 2005. Rosario, Argentina. September 29-30, 2005.
- [5] ISO/IEC 27002. “Information technology - Security techniques - Code of practice for information security controls”. Second edition, 2013-10-01. Reference number ISO/IEC 27002:2013(E)
- [6] J. J. Dujmovic and A. Bayucan, “Evaluation and Comparison of Windowed environments”, Proceedings of the IASTED Interna Conference Software Engineering (SE'97), pp 102-105, 1997.
- [7] J. J. Dujmovic, “A Method for Evaluation and Selection of Complex Hardware and Software Systems”, The 22nd International Conference for the Resource Management and Performance Evaluation of Enterprise Computing Systems. CMG96 Proceedings, vol. 1, pp.368-378, 1996.
- [8] J. J. Dujmovic, “Quantitative Evaluation of Software”, Proceedings of the IASTED International Conference on Software Engineering, edited by M.H. Hamza, pp. 3-7, IASTED/Acta Press, 1997.
- [9] M. Castro, A. Dasso, A. Funes. “Modelo de Evaluación para Sitios de Gobierno Electrónico”. 38 JAIIO/SIE 2009, Simposio de Informática en el Estado 2009, Mar del Plata, Argentina, August 26-28, 2009.
- [10] N. Debnath, A. Dasso, A. Funes, G. Montejano, D. Riesco, R. Uzal, “The LSP Method Applied to Human Resources Evaluation and Selection”, Journal of Computer Science and Information Management, Publication of the Association of Management/International Association of Management, Volume 3, Number 2, 2000, ISBN 1525-4372, pp.1-12.
- [11] N. Debnath, A. Dasso, A. Funes, G.

- Montejano, D. Riesco, R. Uzal, "The LSP Method Applied to Human Resources Evaluation and Selection", Journal of Computer Science and Information Management, Publication of the Association of Management/International Association of Management, Volume 3, Number 2, 2000, ISBN 1525-4372, pp.1-12.
- [12] Narayan Debnath, Aristides Dasso, Ana Funes, Roberto Uzal, José Paganini. "E-government Services Offerings Evaluation Using Continuous Logic". 2007 ACS/IEEE International Conference on Computer Systems and Applications, AICCSA '2007, Amman, Jordan. Sponsored by IEEE Computer Society, Arab Computer Society, and Philadelphia University, Jordan. May 13-16, 2007.
- [13] The Common Criteria <http://www.commoncriteriaportal.org/> (Recuperado marzo 2015)
- [14] The National Information Assurance Partnership/Common Criteria Evaluation and Validation Scheme (NIAP/CCEVS) <https://www.niap-ccevs.org/> (Recuperado marzo 2015)
- [15] Wikipedia http://en.wikipedia.org/wiki/Common_Criteria#Criticisms (Recuperado marzo 2015)
- [16] William Jackson, "Under attack". GCN. 2007. <http://gcn.com/articles/2007/08/10/under-attack.aspx> (Recuperado marzo 2015)
- [17] Barabanov, Rostyslav; Kowalski, Stewart; Yngström, Louise. "Information Security Metrics: State of the Art". DSV Report series No 11, 2011. <http://www.diva-portal.org/smash/get/diva2:469570/FULLTEXT01.pdf> (Recuperado marzo 2015)
- [18] Department of Homeland Security (DHS) Control Systems Security Program (CSSP). Cyber Security Evaluation Tool. https://ics-cert.us-cert.gov/sites/default/files/documents/DHS_CyberSecurity_CSSP-CSET-v4.pdf (Recuperado marzo 2015)
- [19] Payment Card Industry Security Standards Council https://www.pcisecuritystandards.org/security_standards/role_of_pci_council (Recuperado marzo 2015).php
- [20] LeMay, E.; Ford, M.D.; Keefe, K. ; Sanders, W.H. ; Muehrcke, C. "Model-based Security Metrics Using ADversary View Security Evaluation (ADVISE)". Eighth International Conference on Quantitative Evaluation of Systems (QEST), 5-8 Sept. 2011, Publisher: IEEE. Print ISBN: 978-1-4577-0973-9
- [21] Andy Ju An Wang. "Information security models and metrics". ACM-SE 43 Proceedings of the 43rd annual Southeast regional conference – Volume 2, Pages 178-184. ACM New York, NY, USA ©2005 ISBN:1-59593-059-0

Análisis comparativo de Algoritmos Criptográficos Livianos para dispositivos RFID de bajo costo

Mg. Jorge Eterovic; Mg. Domingo Donadello; Esp. Marcelo Cipriano;
Lic. Mara Capuya; Esp. Pablo Pomar

Programa CyTMA2 / Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza
Florencio Varela 1903 (B1754JEC), San Justo, (5411) 4480-8900

eterovic@unlam.edu.ar; ddonadel@ing.unlam.edu.ar; cipriano1.618@gmail.com;
mcapuya@gmail.com; pablo_pomar@yahoo.com.ar

1. Resumen.

El desarrollo de la Internet de los objetos dará lugar a un inmenso despliegue de millones de objetos inteligentes que interactuarán entre sí y con Internet. El papel de la tecnología RFID será primordial en este escenario.

El ritmo de adopción del RFID es vertiginoso y ya se ha convertido en una realidad. Las etiquetas de bajo costo representan el mayor desafío en términos de seguridad y privacidad pero sus escasos recursos influyen sobre los métodos criptográficos existentes.

El objetivo de éste proyecto de investigación es realizar un análisis comparativo del comportamiento de los Algoritmos Criptográficos Livianos existentes para ser usados en dispositivos RFID de bajo costo.

Palabras Clave:

Criptografía Ligera, Algoritmos Criptográficos Livianos, RFID

2. Contexto.

La Universidad Nacional de La Matanza mantiene una política en la que se fomenta y promueve la investigación académica y la inclusión en ella de alumnos de grado, posgrado y maestría.

Es por ello que esta línea de investigación se enmarca en el siguiente programa:

- Programa CyTMA2 (Programa de Investigación Científica, Desarrollo y Transferencia de Tecnologías e Innovaciones. UNLaM).

3. Introducción.

Cuando se habla sobre Internet de las Cosas, Internet of Things (IoT por sus siglas en inglés), en realidad de lo que se está hablando es de la conectividad a través de Internet entre objetos. Pero Internet de las cosas va mucho más allá. Estas cosas van desde electrodomésticos controlados por un Smartphone hasta niveles más profesionales.

Con la Internet de las Cosas, todo lo real se convierte en virtual, lo que significa que cada persona y las cosas tienen una ubicación en Internet. Estas entidades virtuales pueden producir y consumir

servicios y colaborar entre sí con un objetivo en común.

La manera en que estos objetos pueden comunicar o recibir información es a través de sensores que en algunos casos, pueden visualizarse. Pero no siempre es posible notar su presencia. Dentro de la conexión de los objetos con los sistemas de información, dos son las tecnologías clave que ya se están insertando en diversos sectores de la industria para acercar la Internet de las Cosas a la realidad. Estas tecnologías son la identificación por radiofrecuencia (RFID) y las redes de sensores inalámbricas.

La International Telecommunication Union (ITU), en su "Informe sobre la Internet de las Cosas" califica a la tecnología RFID como un "pivote que habilitará el Internet de las Cosas", permitiendo la conversión de los "objetos cotidianos" en "inteligentes" [1]. Sin embargo, sin bases sólidas de seguridad, es posible que estos objetos sean pasibles de ataques. Estas amenazas podrían llegar a ser cada vez más perjudiciales que cualquiera de sus beneficios [2,3].

Las investigaciones sobre algoritmos criptográficos están avanzadas y cada día se generan nuevos algoritmos para las claves de autenticación. La investigación académica y la Asociación Internacional de Investigaciones en Criptografía (IACR-International Association for Cryptologic Research) [4], en particular, impulsaron la definición de distintos mecanismos que proporcionaron un nivel de seguridad y privacidad adecuados a las limitaciones del hardware de las etiquetas RFID.

Se trabaja continuamente sobre el área que dio en llamarse "Criptografía Ligera" [5] y se abordan los temas de privacidad, protección de datos personales y seguridad en las comunicaciones electrónicas

sobre las amenazas específicas para las aplicaciones RFID.

La Criptografía Ligera o Liviana (LICRYPT - Lightweight Cryptography) es un nuevo campo de investigación que apunta a estudiar métodos criptográficos con el fin que puedan utilizarse en objetos inteligentes.

La LICRYPT está orientada a Hardware o Software, determinándose parámetros para evaluar y medir las implementaciones que apliquen a este tipo de criptografía. Por ejemplo para hardware se estudian el tamaño de los chips y el consumo de energía que se requiere. Para software, en cambio, se analizan la longitud del código, el uso y consumo de memoria Ram.

En LICRYPT se podrán encontrar: algoritmos de clave pública, clave privada, block ciphers y stream ciphers. Además funciones hash y mecanismos de autenticación, como pueden hallarse en la criptografía tradicional.

La comunidad científica, a la actualidad, aún no tiene un criterio determinado para clasificar a un algoritmo criptográfico como ligero. Lo que sí está claro es que las técnicas criptográficas involucradas tienen que usar la mínima cantidad de recursos posibles de los objetos en los que se las aplicará.

4. Líneas de Investigación, Desarrollo e Innovación.

Muchos son los avances a nivel de criptografía que se están realizando, pero no todos los algoritmos livianos son eficientes a la hora de implementarlos en la seguridad de los RFID de bajo costo.

La presunción que es posible evaluar el desempeño general de un algoritmo perteneciente a LICRYPT funcionando

sobre una determinada plataforma móvil [6,7,8].

El mismo podrá ser un teléfono celular en particular, una tablet, un dron, lentes como el google glass o equivalentes con capacidad de comunicaciones, o cualquier otro tipo de dispositivo inteligente y portable, para el cual querrá hacer que ejecute un algoritmo determinado. Tal equipo será simulado a través de una virtualización.

De esa manera obtener las métricas que permitan determinar la performance del mismo y emitir un juicio acerca de su comportamiento.

Esta línea de investigación propone evaluar la posibilidad de determinar el funcionamiento performático de un algoritmo criptográfico [9-16] ejecutado en distintos perfiles de HW/SW. Luego, poder evaluar su comportamiento para ser aplicados en dispositivos RFID de bajo costo.

5. Resultados y Objetivos.

El objetivo de este proyecto es realizar un análisis comparativo, de acuerdo a criterios de aplicabilidad y seguridad, de 3 Algoritmos Criptográficos Livianos para dispositivos RFID de bajo costo.

Se realizará un relevamiento exhaustivo de los principales algoritmos criptográficos ligeros existentes y determinará cuáles se podrían utilizar para dispositivos RFID de bajo costo.

Se definirán indicadores utilizando otras experiencias internacionales para evaluar comportamientos y permitir comparaciones.

Se simulará el funcionamiento de los algoritmos seleccionados y se realizará una tabla comparativa sobre el comportamiento de los algoritmos estudiados.

Finalmente se redactará un informe final y se presentarán en diferentes congresos los resultados obtenidos de esta investigación, para difusión y conocimiento de la comunidad científica.

Se desarrollará un capítulo específico de “Criptografía Ligera” en la materia electiva Criptografía y “Aplicaciones de Criptografía Ligera” en la materia Auditoria y Seguridad informática de la carrera de Ingeniería en Informática del DIIT.

6. Formación de Recursos Humanos.

La Lic. Mara Capuya se suma al equipo de investigadores como alumna de la Maestría en Informática de la UNLaM. Tanto la Lic. Capuya como el Esp. Pablo Pomar se encuentran desarrollando su trabajo de tesis de posgrado de la Maestría en Informática. Ambos están siendo tutorados por el Mag. Jorge Eterovic, director del proyecto de investigación y por el Esp. Marcelo Cipriano.

Asimismo parte del equipo de investigación dictan la asignatura electiva Criptografía en el 5to. Año de la carrera de Ingeniería Informática de la UNLaM, invitarán a sus alumnos a participar de la investigación. Dado que es un proyecto nuevo, aún no se ha logrado la incorporación de ningún alumno.

7. Referencias

[1] Instituto Nacional de Tecnologías de la Comunicación y Agencia Española de Protección de Datos.: Guía sobre seguridad y privacidad de la tecnología RFID. Spain. 2010. www.inteco.es

[2] Román R., Nájera P., López J. ”Securing the Internet of Things”. University of Malaga, Spain. 2011.

- [3] Román R., Nájera P., López J. “Los Desafíos De Seguridad En La Internet De Los Objetos” University of Malaga, España. 2010.
- [4] International Association for Cryptologic Research. 2015. <http://www.iacr.org/events/>
- [5] Bhattasali Tapalina. “LICRYPT: Lightweight Cryptography Technique for Securing Smart Objects in Internet of Things Environment”. Research Scholar, University of Calcutta. 2013.
- [6] Heer, T.; Garcia-Morchon, O.; Hummen, R.; Keoh, S.L.; Kumar, S.S.; Wehrle, K. “Security challenges in the IP-based internet of things”. *Wirel. Pers. Commun.* 61, 527–542. 2011.
- [7] Garcia-Morchon, O.; Keoh, S.; Kumar, S.; Hummen, R.; Struik, R. “Security Considerations in the IP-based Internet of Things”. IETF Internet Draft draft-garcia-core-security-04; The Internet Engineering Task Force (IETF): Fremont, CA, USA, 2012.
- [8] Cirani S., Ferrari G., Veltri L. “Enforcing Security Mechanisms in the IP-Based Internet of Things: An Algorithmic Overview”. *Algorithms* 2013, 6, 197-226;
- [9] Suzaki, T., Minematsu K., Morioka S., Kobayashi E. “TWINE: A Lightweight, Versatile Block Cipher”. NEC Corporation, Japan. 2014.
- [10] Bogdanov A., Knudsen L., Leander G. et al. “PRESENT: An Ultra weight Block Cipher”. Springer-Verlag Berlin Heidelberg 2007 HES 2007, LNCS 4727, pp. 450–466. 2007.
- [11] Wentao Z. , Zhenzhen B., Dongdai L., Rijmen V., Yang B., Verbauwhede I. “RECTANGLE: A Bit-slice Ultra-Lightweight Block Cipher Suitable for Multiple Platforms”. China, Bélgica. 2015.
- [12] Beaulieu R., Shors D y otros. “The SIMON and SPECK Families of Lightweight Block Ciphers”. *Cryptology ePrint Archive: Report 2013/404*. 2013.
- [13] Mouha N., Mennink B., y otros. “Chaskey: An Efficient MAC Algorithm for 32-bit Microcontrollers”. Department of Electrical Engineering, Leuven and iMinds, Bélgica. 2013.
- [14] Hongjun W., Tao H. “JAMBU Lightweight Authenticated Encryption Mode and AES-JAMBU (v1)” Division of Mathematical Sciences Nanyang Technological University, Singapur. 2014.
- [15] Engels D., Fan X., Gong G., Hu , H , Smith M. “Hummingbird: Ultra Lightweight Cryptography for Resource-Constrained Devices.” 1st International Workshop on Lightweight Cryptography for Resource-Constrained Devices (WLC’2010). Tenerife, Canary Islands, Spain, 2010.
- [16] ISO/IEC 29192-2:2012. Information technology - Security techniques - Lightweight cryptography - Part 2: Block ciphers. 2012.

Anonimato en Sistemas de e-Voting: Últimos Avances

Pablo García¹; Germán Montejano^{1 2}; Silvia Bast¹; Estela Fritz¹

¹ Departamento de Matemática
Universidad Nacional de La Pampa
Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina
Tel.: +54-2954-245220– Int. 7125
[pablogarcia, silviabast, fritzem]@exactas.unlpam.edu.ar

² Departamento de Informática
Universidad Nacional de San Luis
Ejército de los Andes 950 – (5700) San Luis – San Luis – Argentina
Tel.: +54-266-4520300– Int. 2128
gmonte@unsl.edu.ar – web: <http://www.unsl.edu.ar>

Resumen

Desde el año 2012 se trabaja en la investigación relacionada con la protección de la privacidad de los votantes en un sistema de e-Voting. En ese sentido, se plantea que debe ser de mayor nivel la protección del anonimato que la que se le otorgue al proceso electoral, dado que este último debe ser protegido por un período finito de tiempo, mientras que el anonimato debe asegurarse indefinidamente.

En consecuencia, se considera probado que el nivel de seguridad para el anonimato debe ser incondicional, es decir que será seguro aún cuando un criptoanalista cuente con tiempo y recursos ilimitados. En cambio, puede aceptarse que al proceso de elección se le otorgue un nivel de seguridad computacional de razonable magnitud, teniendo en cuenta que luego de unas pocas horas, los resultados serán conocidos públicamente.

Fundamentalmente, los avances se relacionan con optimizaciones obtenidas sobre el protocolo Non-Interactive Dining Cryptographers (NIDC, [1]) el cual es un derivado de [2], que incorpora características asíncronas.

En el presente documento se exponen los avances realizados en los últimos doce meses en el ámbito de la protección del anonimato y se enuncian las acciones futuras a desarrollar.

Palabras clave: *Anonimato, e-Voting, Non-Interactive Dining Cryptographers, Seguridad Incondicional.*

Contexto

Por Resolución N° 488/14 del Consejo Directivo de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa se acredita el Proyecto de Investigación: "Aspectos de Seguridad en Proyectos de Software", que se desarrolla en el ámbito de la FCEyN de la UNLPam. El mismo es dirigido por el Doctor Germán Antonio Montejano y codirigido por el Magister Pablo Marcelo García e incluye a la Licenciada Silvia Gabriela Bast y la Profesora Estela Marisa Fritz como investigadoras.

El Proyecto surge desde la línea de Investigación "Ingeniería de Software y Defensa Cibernética", presentada en [3], y que a su vez se enmarca en el Proyecto "Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la

Profesión de Ingeniero de Software” de la Facultad de Ciencias Físico - Matemáticas y Naturales de la Universidad Nacional de San Luis (UNSL) (<http://www.sel.unsl.edu.ar/pro/proyec/2012/index.html>) y que incluye acciones de cooperación con la Universidad Federal de Minas Gerais (UFMG, Brasil). Entre tales acciones debe mencionarse que Jeroen van de Graaf, PhD., Docente de UFMG, y el Dr. Germán Montejano (UNSL) fueron orientadores del Mg. Pablo García en el desarrollo de su tesis de maestría titulada “Optimización de un Protocolo Dining Cryptographers Asíncrono”, defendida en 2013. Durante el desarrollo de la misma se generaron una serie de publicaciones de avances parciales, como por ejemplo [4], [5], [6], [7] y [8].

Introducción

El voto electrónico como alternativa al método manual tradicional presenta partidarios y detractores, en proporciones similares. Ambas posturas proponen argumentos razonables.

En consecuencia, la implementación de sistemas de voto electrónico exige que el escrutinio asociado refleje de manera indiscutible la voluntad de los ciudadanos y que, simultáneamente, los electores vean garantizada su privacidad de manera indefinida.

En particular, desde este proyecto se presta máxima atención a las consecuencias que puede acarrear, para cualquier votante, el hecho de que su voto se conozca. Fundamentalmente, a las prácticas deshonestas que se derivan de conocimiento de esa información. Tales prácticas pueden producirse con o sin el aval del elector. Por ejemplo, si un ciudadano pudiera probar que votó a un determinado partido político, podría obtener una contraprestación. Del mismo modo, si un sector detecta que un votante votó a otra opción, podría llevar

a cabo acciones que perjudiquen al mismo.

Como consecuencia de lo anterior, la investigación otorga máximo interés a los protocolos que garanticen el anonimato incondicional y no exijan la concurrencia online de la totalidad de los participantes. NIDC cumple con ambos requisitos; es por este motivo que se trabajó en la optimización de algunos aspectos de la propuesta original:

- Nuevo protocolo antifraudes, basado en logaritmos discretos y commitments de Pedersen, que se presenta en [4] y que mantiene el nivel de seguridad original, (basado en BCX), con una mayor eficiencia en el uso de los recursos
- Esquema alternativo de almacenamiento de sufragios basado en canales paralelos de slots, que exige menor cantidad de almacenamiento ofreciendo para otorgar un nivel de seguridad determinado. Este esquema se presenta originalmente en [7].
- Implementación de múltiples redes NIDC en serie o paralelo con fines similares al párrafo anterior ([8]).

Líneas de Investigación, Desarrollo e Innovación

El grupo de trabajo investiga, básicamente sobre tres campos relacionados:

- Protección del anonimato de los votantes en sistemas de voto electrónico ([9]).
- Integridad de los datos de un sistema de e-Voting ([10]).
- Integridad de las bases de datos pertenecientes a un sistema de gestión de aprendizaje ([11]).

Resultados y Objetivos

En el ámbito de la protección de la privacidad, este grupo de trabajo ha realizado las siguientes publicaciones:

- [12]: Presenta una técnica basada en almacenamiento de sufragios basada en canales paralelos de slots, exponiendo una serie de fórmulas matemáticas que describen el comportamiento del mismo.
- [13]: Este documento expone un método sistemático de elección de los parámetros óptimos para la implementación de un esquema del tipo descripto en el párrafo anterior.
- [14]: Muestra una generalización del enfoque basado en canales paralelos y muestra los resultados obtenidos con la implementación de un simulador de actos electorarios.

A futuro, se pretende llevar a cabo las siguientes acciones:

- Implementación de una aplicación experimental que permita observar el comportamiento de los modelos propuestos en las publicaciones producidas, tanto en lo referido a la protección de la privacidad como a la seguridad de las bases de datos relacionadas con un sistema de voto electrónico.
- Continuar con el relevamiento de aplicaciones orientadas al voto electrónico, con el fin de detectar fallencias y proponer mejoras.

- Ampliar el simulador de actos electorarios para su generalización y publicación online.

Formación de Recursos Humanos

En el marco del presente proyecto se presentan los siguientes puntos relacionados con la formación de recursos humanos:

- Pablo García realizó una estadía de un año en la Universidad Federal de Minas Gerais (UFMG), aprobando seminarios de posgrado y trabajando en el grupo “Criptografía Teórica y Aplicada”, dirigido por Jeroen van de Graaf, PhD.
- Pablo García defendió su tesis para obtener el grado de Magister en Ingeniería de Software de la Universidad Nacional de San Luis, bajo la dirección de Jeroen van de Graaf, PhD (UFMG) y Dr. Germán Montejano (UNSL). La tesis se tituló: “Optimización de un Esquema Dining Cryptographers Asíncrono” y recibió la calificación de sobresaliente.
- Silvia Bast está desarrollando su tesis para obtener el grado de “Especialista en Ingeniería de Software”. Su plan de trabajo fue aprobado y se planea su defensa para mayo de 2016. La tesis se titula: “Sistemas de E-Voting: Integridad de Datos” y está dirigida por el Dr. Germán Montejano (UNSL) y el Magister Pablo García (UNLPam).
- Pablo García está desarrollando su tesis para obtener el grado de “Especialista en Ingeniería de Software”. Su plan de trabajo fue aprobado y se planea su

defensa para septiembre de 2015. La tesis se titula: "Anonimato en sistemas de Voto Electrónico" y es dirigida por Jeroen van de Graaf, PhD (UFMG) y Dr. Germán Montejano (UNSL).

- Silvia Bast y Pablo García completaron el cursado de la totalidad de los créditos exigidos en el Doctorado en Ingeniería Informática en la Facultad de Ciencias Físico Matemáticas y Naturales de la Universidad Nacional de San Luis (UNSL).
- Estela Fritz está desarrollando su tesis para obtener el grado de "Especialista en Tecnologías Informáticas aplicadas en Educación". Su plan de trabajo fue aprobado y se planea su defensa para octubre de 2016. La tesis se titula "Propuesta de clasificación de software libre utilizado en la enseñanza de la programación" y es dirigida por Mg. Alejandra Zangara (UNLP).

Referencias

[1] van de Graaf J.: "Anonymous One Time Broadcast Using Non Interactive Dining cryptographer Nets with Applications to Voting". Publicado en: "Towards trustworthy Elections". Ps 231-241. Springer-Verlag Berlin, Heidelberg. ISBN:978-3-642-12979-7. 2010.

[2] Chaum D.: "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability". Journal of Cryptology. 1988.

[3] Uzal R., van de Graaf J., Montejano G., Riesco D., García P.: "Inicio de la Línea de Investigación "Ingeniería de Software y Defensa Cibernética". Memorias del XV

Workshop de Investigadores en Ciencias de la Computación 2013 (WICC 2013). Ps.769 - 773. ISBN: 9789872817961. 2013.

[4] van de Graaf J., Montejano G., García P.: "Optimización de un Protocolo Non-Interactive Dining Cryptographers". Congreso Nacional de Ingeniería Informática / Sistemas de Información . CoNaIISI 2013. 21 y 22 de noviembre de 2013. Córdoba, Argentina. NACIONAL.

[5] van de Graaf J., Montejano G., García P., Bast S.: "Anonimato en Sistemas de Voto Electrónico". Memorias del XVI Workshop de Investigadores en Ciencias de la Computación 2014 (WICC 2014). Ps. 822 - 826. ISBN: 9789503410844 . 8 y 9 de mayo de 2014. NACIONAL.

[6]. van de Graaf J., Montejano G., García P.: "Manejo de Colisiones en un Protocolo Non Interactive Dining Cryptographers". Anales de las 42° Jornadas Argentinas de Informática e Investigación Operativa (JAIIO, ISSN: 1850-2776). Workshop de Seguridad Informática (WSegI 2013, ISSN: 2313-9110). Ps. 29 a 43. Septiembre 2013. NACIONAL.

[7] García P., van de Graaf J., Montejano G., Bast S., Testa O.: "Implementación de Canales Paralelos en un Protocolo Non Interactive Dining Cryptographers". 43° Jornadas Argentinas de Informática e Investigación Operativa (JAIIO 2014), Workshop de Seguridad Informática (WSegI 2014). Trabajo en aceptado para su presentación. NACIONAL.

[8] García P., van de Graaf J., Hevia A., Viola A.: "Beating the Birthday Paradox in Dining Cryptographer Networks". The third International Conference on Cryptology and Information Security in Latin America, Latincrypt 2014. September 17-19, 2014. Florianopolis, Brasil. Lecture Notes in Computer Science, Springer (2014). INTERNACIONAL.

[9] García P., Montejano G., Bast S., Fritz E.: "Seguridad Incondicional para el Anonimato en Sistemas de e-Voting". XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015). 16 y 17 de abril de 2015. Facultad de Ciencias Exactas. Universidad Nacional de Salta. ISBN: 978-987-633-134-0. NACIONAL.

[10]. Bast S., Montejano G., García P., Fritz E.: "Evaluación de la integridad de datos en Sistemas de e-Voting". XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015). 16 y 17 de abril de 2015. Facultad de Ciencias Exactas. Universidad Nacional de Salta. ISBN: 978-987-633-134-0. NACIONAL.

[11] Fritz E., Montejano G., García P., Bast S.,: "Integridad de Datos en Sistemas de Gestión de Aprendizaje". XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015). 16 y 17 de abril de 2015. Facultad de Ciencias Exactas. Universidad Nacional de Salta. ISBN: 978-987-633-134-0. NACIONAL.

[12] García P., van de Graaf J., Montejano G., Riesco d, Debnath N., Bast S.: "Storage Optimization for Non Interactive Dining Cryptographers (NIDC)". 12th International Conference on Information Technology : New Generations (ITNG 2015). April 13-15, 2015, Las Vegas, Nevada, USA.

[13] García P., van de Graaf J., Montejano G., Riesco D., Debnath N., Bast S.: "A Systematic Method for Choosing Optimal Parameters for Storage in Parallel Channels of Slots". 2016 IEEE International Conference on Industrial Technology (ICIT2016). 14-17 de marzo de 2016. Taipei, Taiwan. Trabajo aceptado para su publicación. Previamente el trabajo fue aceptado en NICS 2015, AICCSA 2015 y 12th ACS/IEEE International Conference on Computer Systems and Applications, pero no fue publicado por falta de financiación.

[14] García P., Montejano G. Bast S., Fritz E., Riesco D., Debnath N.,: "A Proposal for Anonymous Data Storage". IEEE International Conference on Industrial Informatics. INDIN 2016. 18 al 21 Julio de 2016, Futuroscope-Poitiers, Francia. Trabajo enviado para su evaluación.

Esteganografía: Sustitución LSB 1 bit utilizando Matlab

Mg. Ing. Gustavo Rodríguez Medina¹, Mg. Ing. G. Sergio Navas²

Gabinete de Computación / Fac. de Ingeniería / Univ. Nacional de San Juan^{1,2}

Av. Libertador Gral. San Martín 1109 (oeste) – San Juan
0264 – 4211700 (Int. 285¹ / 435²)

grodriguez@unsj.edu.ar¹, snavas@unsj.edu.ar²

Resumen

Existen diferentes técnicas a través de las cuales se puede ocultar información (mensaje digital) en un portador, con el fin de ser enviado a un receptor, y que el hecho pase inadvertido para terceros.

Los portadores preferidos para implementar esteganografía son los de tipo multimedia (video, imagen, sonido). El presente tratado se enfoca particularmente en portadores imagen BMP de 24bpp.

Entre las técnicas se encuentra la de Sustitución de bits, y a su vez, la más conocida es la de sustitución del bit menos significativo, cuya sigla se conoce como "LSB 1 bit"¹.

Si bien, esta técnica puede implementarse desarrollando un programa de computación en cualquier lenguaje de programación, se ha encontrado que utilizando el Software matemático Matlab, y a través de su lenguaje de programación, se obtienen algunas ventajas con respecto al resto. Estas ventajas están dadas por la forma en que Matlab opera con las imágenes (matricialmente).

En este trabajo se expone la implementación de la técnica esteganográfica de Sustitución LSB 1 bits desarrollado en Matlab, haciendo uso de herramientas disponibles para el procesamiento de

imágenes. El que resulta ser de bastante utilidad a los fines de desarrollar una herramienta de Esteganografía aplicada.

Palabras clave:

Esteganografía, Sustitución LSB 1 bit, Implementación en Matlab.

Contexto

El presente trabajo expone de manera resumida uno de los trabajos abordados en el marco de la tesis "*Estudio, análisis, desarrollo y propuestas de algoritmos para la selección óptima de métodos de sustitución en aplicaciones esteganográficas*", para la Maestría en Informática de la Universidad Nacional de la Matanza, bajo la dirección de docentes/investigadores de esa Universidad y de la Universidad Nacional de San Juan.

Introducción

Un Portador imagen puede ser utilizado para ocultar, a la vista de intrusos, cualquier mensaje u objeto software (archivo), codificándolo como sutiles cambios en los colores de los píxeles (sus componentes RGB) que no deben ser percibidos por el ojo humano; de tal forma que el Portador que contiene el mensaje,

¹ LSB: *Least Significant Bit*

o "Estegoportador", pueda ser transmitido, sin que sea detectado el hecho, para luego aplicar el proceso inverso (decodificación) de modo que el receptor pueda recuperar y disponer del Mensaje enviado [1].

En general, cualquier archivo multimedia contiene áreas de datos poco significativas. Esas áreas se pueden sustituir por otros datos, realizando cambios que son imperceptibles a la claridad visual (o auditiva) humana [2].

Esto permite encubrir información de interés dentro de un archivo Portador, haciendo que el mismo parezca igual al original. El método² *LSB 1 bit* consiste en alterar el **bit menos significativo** de los bytes componentes de color del Portador, siendo ésta la razón de la denominación de la técnica "*Least Significant Bit*". La teoría indica que con la sustitución del bit "menos significativo" el ojo y oído, según el portador, no serán capaces de detectar los cambios.

En general los métodos de sustitución, y en particular LSB, son más efectivos si los archivos portadores tienen mucho "ruido", en el sentido de que las imágenes tengan bruscos cambios de color entre pixeles adyacentes, siendo lo opuesto a tener imágenes con amplias áreas de color uniforme [2].

Una característica importante, es que el método de sustitución LSB no incrementa el tamaño del archivo portador [3], y dependiendo del porte del Mensaje a ocultar, puede causar cambios significativos en el espectro de color del Portador respecto a la versión original, lo que puede no ser perceptible pero si detectable bajo análisis espectral y/o estadístico [2].

En el presente trabajo se opera con imágenes bitmap BMP de 24 bpp de pro-

fundidad de color, para el caso del Portador.

Forma habitual de implementar LSB 1 bit:

Para entender mejor la forma general en la cual se implementa la técnica base de sustitución 1 bit, se debe observar la figura 1. En ésta, se aprecia que de cada byte del mensaje se extrae cada bit, y éste a su vez es insertado en cada bit menos significativo de cada componente de color RGB, o en cada byte, que conforma a cada uno de los pixel de la imagen portadora. La figura 1 hace referencia a pixeles de una imagen portadora BMP de 24 bits.

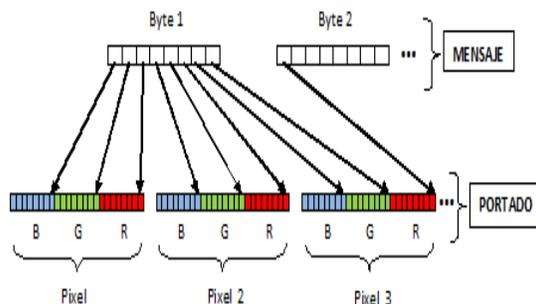


Figura 1. Implementación habitual LSB 1 bit.

Es importante tener en cuenta que cuando se opera sobre una imagen desde su archivo binario, el orden de los bytes que representa a las componentes R, G y B de cada pixel, presentan el orden B, G y R. Ese orden es el que se muestra en la figura 1 por tratarse de la descripción genérica del esquema de implementación de la técnica, y es la forma en que se encuentra descrita en la mayor parte de la literatura y antecedentes recopilados. Corresponde con el orden que presenta cada uno de las componentes de color cuando se lee una imagen byte a byte desde su almacenamiento en disco. Siendo también el mismo orden de canales (B, G y R) en

² En este trabajo se hace mención a Método y a Técnica de manera indistinta.

el que se opera cuando se manipula la imagen desde un programa escrito en cualquiera de los lenguajes habituales de programación.

Manipulación de imágenes como matrices en Matlab:

Matlab es un software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio (lenguaje M). Está disponible para las plataformas Unix, Windows y Apple Mac OS X. [4].

Está orientado para llevar a cabo computaciones numéricas casi de todo tipo, pudiendo manipular vectores y matrices tanto reales como complejas con funciones y fórmulas de variadas ramas de la matemática. Matlab se compone de un programa básico y un conjunto de *Toolbox* para labores más especializadas [4].

La ventaja de operar imágenes con Matlab, es que las mismas se cargan al entorno de trabajo en forma de matrices, y dado que este software está orientado a aplicar toda su potencialidad de cálculo a las matrices, es que resulta una herramienta poderosa para la manipulación y procesamiento matemático; además de poder procesarse digitalmente las imágenes a través del *Toolbox de Procesamiento de imágenes* con el que cuenta (por ej. aplicar filtros, mascarar, etc.).

Para el presente trabajo, las imágenes de 24 bpp. son representadas en Matlab por una matriz tridimensional de $m \times n \times p$, donde m y n representan el alto y el ancho de la imagen respectivamente, y p el plano de color RGB. En Matlab el plano p se asocia con el valor "1", para el plano o canal de color Rojo, el valor "2" con el color Verde y el valor "3" con el color Azul.

En la figura 2 se observa la descomposición de una imagen de 24 bpp en 3 planos de color.

También, indica que una imagen que se lee desde Matlab y se carga en su

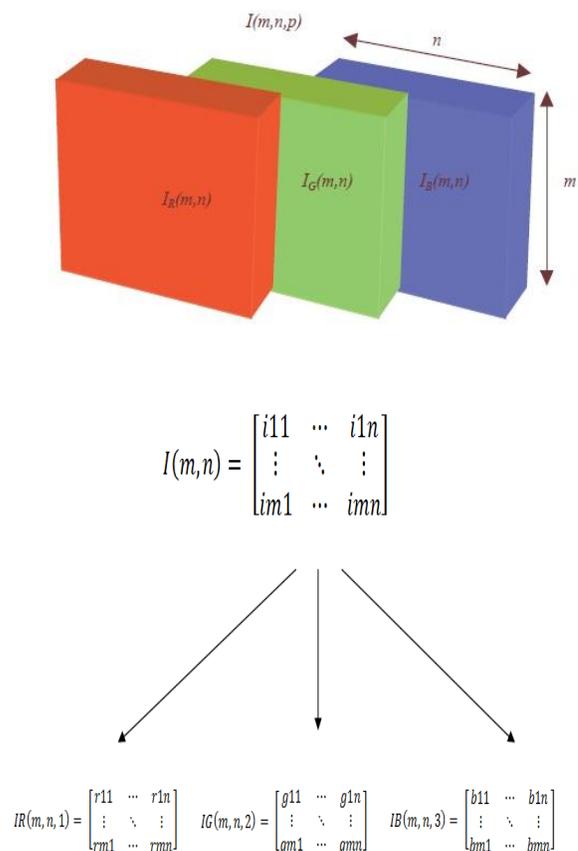


Figura 2. Operatoria de imágenes en Matlab

entorno de trabajo, genera una matriz a la cual se le ha denominado "I", la que a su vez se descompone en 3 matrices, IR, IG e IB, cada una de éstas corresponde a los planos de color R, G y B, respectivamente.

Implementación de LSB 1 bit en Matlab:

Según lo expuesto en "Formo habitual de implementar LSB 1 bit", se vio que en el caso de trabajar a las imágenes en forma *No Matricial* (forma habitual que se expone la técnica en la mayoría de los antecedentes estudiados y que se opera con las imágenes cuando no se utiliza Matlab), el orden de dichos canales de color, que componen los pixeles es B, G y R. El esquema de implementación general

de la técnica LSB 1 bit, mostrado en la figura 1 responde justamente a la operatoria con imágenes de forma No Matricial. Por lo cual para adecuarlo al presente trabajo y poder realizar la implementación algorítmica en Matlab, se hace una pequeña modificación en dicho esquema, resultando el de la figura 3.

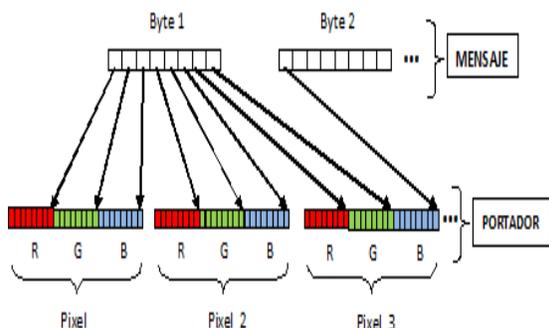


Figura 3. Implementación LSB 1 bit en Matlab.

La diferencia presentada entre las figuras 1 y la figura 3, si bien gráficamente exponen simplemente un cambio en el orden de las componentes R, G y B que constituyen cada píxel, deben ser tenidas en cuenta para la implementación algorítmica de la técnica, como así también en el proceso inverso de recuperación o de decodificación del Mensaje oculto en el Portador.

Se aclara que tal modificación no representa una redefinición de la técnica, sino una adaptación que permita la implementación en Matlab en el marco del presente trabajo.

Líneas de Investigación, Desarrollo e Innovación

La línea de investigación corresponde a la temática de *Esteganografía*, la que se enmarca en el área de Seguridad Informática. En este sentido, los autores del presente trabajo han elaborado dos tesis de Maestría, como así también la publica-

ción de resultados en diferentes congresos.

Los resultados obtenidos desde el año 2006, aportan estudios y desarrollos innovadores en el área.

Cabe destacar que en el mes de octubre de 2015, fue defendido el trabajo de tesis titulado "*Estudio, análisis, desarrollo y propuestas de algoritmos para la selección óptima de métodos de sustitución en aplicaciones esteganográficas*".

Resultados y Objetivos

Se ha desarrollado en Matlab una función denominada *LSB1()*, a través de un archivo ".m", que implementa el algoritmo de la técnica de Sustitución LSB 1 bit, operando a la imagen Portadora como una matriz $m \times n$, siendo m y n el alto y ancho respectivamente de la imagen BMP de 24 bpp.

El llamado de dicha función, desde la línea de comandos de Matlab, como primer paso solicita al usuario que busque y seleccione (a través de una ventana emergente) a la imagen portadora elegida (la que debe ser de formato BMP de 24 bpp.) como así también al archivo Mensaje (el que puede ser de cualquier tipo). Luego la función desarrollada aplica el algoritmo de sustitución LSB 1 bit implementado, generando como resultado la presentación gráfica por pantalla de la imagen Portadora seleccionada, como también a la imagen Estegoportadora.

A modo de ejemplo se presenta el resultado generado por el algoritmo implementado en la función *LSB1()*. Se elige como Portador a la imagen "*Mariposa.bmp*" de 712x475 píxeles, y como Mensaje al archivo imagen "*atardecer.jpg*" de 36.151 bytes. Este último no se expone en este trabajo debido a la limitación respecto a la cantidad de hojas permitidas para el artículo.

La visualización generada de ambas imágenes durante la ejecución de la función, es a los fines de poder comparar visualmente y analizar si se percibe alguna diferencia entre ellas. En este trabajo no se expone la imagen Portadora, debido que para esta técnica entre la imagen Portadora y la Estegoportadora no presentan efectos visuales perceptibles a simple vista (resultado deseado).



Figura 4. Estegoportador obtenido.

En la figura 4 se presenta a la imagen Estegoporador, de la cual no se advierte ninguna marca o diferencia con respecto a la imagen original (portador), lo cual es el objetivo buscado en la esteganografía orientada a comunicaciones encubiertas.

En la figura 5 se ha aplicado un filtro y realce al efecto esteganográfico de la imagen con el fin de poder visualizar los pixeles afectados para observar el trabajo de sustitución (a modo demostrativo).

Se puede apreciar en la figura 5, en su parte superior, una franja con pixeles coloreados, los que representan a la sustitución del bit menos significativo de cada plano de color correspondiente al Portador, por todos los bits que forman al Mensaje. Se debe recordar que este efecto es exagerado a los fines demostrativos del ejemplo.

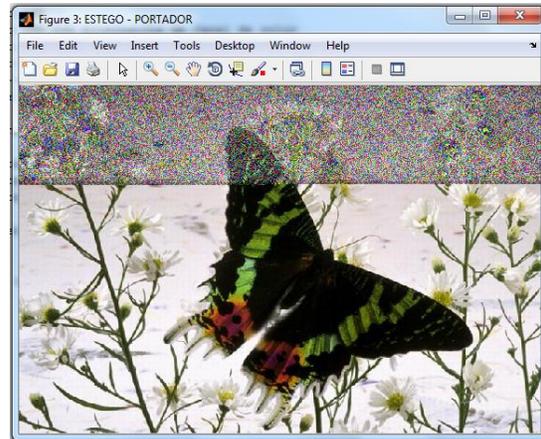


Figura 5. Estegoportador con filtro y realce del efecto de sustitución.

El algoritmo implementado en esta función de Matlab ha sido testado a través de numerosas pruebas, obteniendo excelentes resultados. Cabe destacar que este algoritmo ha servido como base para el desarrollo e implementación de variantes a esta técnica. También ha sido ampliamente utilizado durante el desarrollo de la Tesis de Maestría en Informática del autor principal del presente trabajo.

No se expone el algoritmo por el limite de espacio permitido para el artículo.

Formación de Recursos Humanos

En la temática de Esteganografía se viene trabajando desde el año 2005.

Durante el año 2006, el segundo autor de este trabajo defendió su Tesis de Maestría denominada “*Exploración de efectos esteganográficos sobre portadores imagen de mapa de bits utilizando diferentes técnicas y algoritmos*”.

En octubre de 2015, el autor principal del presente trabajo defendió su Tesis de Maestría titulada “*Estudio, análisis, desarrollo y propuestas de algoritmos para la selección óptima de métodos de sustitución en aplicaciones esteganográficas*”.

Referencias

[1] C. Gustavo Rodriguez M. Estudio, análisis, desarrollo y propuestas de algoritmos para la selección óptima de métodos de sustitución en aplicaciones esteganográficas. Argentina. Univ. Nacional de la Matanza – Escuela de Posgrado. 2015.

[2] G. S. Navas. Exploración de efectos Esteganográficos sobre portadores imagen de mapa de bits utilizando diferentes técnicas y algoritmos. Argentina. Univ. Nacional de la Matanza – Escuela de Posgrado. 2006.

[3] Abbas Cheddad, Joan Condell, Kevin Curran, Paul McKeivitt. Digital image steganography: Survey and analysis of current methods [en línea]. El Sevier (Ed), Journal of Signal Processing, 90(3), 727-752, marzo 2010. [Citado 12 Junio 2014].

Disponible en World Wide Web:
<http://www.sciencedirect.com/science/article/pii/S0165168409003648>

[4] Mathworks. Ayuda en línea de Matlab [en línea]. Estados Unidos.. 1994 - 2015.
Disponible en World Wide Web:
<http://www.mathworks.com/>

Secuencias Seudoaleatorias para Criptología

Antonio Castro Lechtaler^{1,2}; Marcelo Cipriano¹; Edith García¹; Julio Liporace¹
Ariel Maiorano¹, Eduardo Malvacio¹, Néstor Tapia¹, Dulio, Nicolás¹; Pérez, Pablo¹

¹Laboratorio de Investigación en Técnicas Criptográficas y Seguridad Teleinformática.
Escuela Superior Técnica, Facultad de Ingeniería. Instituto Universitario del Ejército.

²UNDeC Universidad Nacional de Chilecito.

acastro@est.iue.edu.ar; marcelocipriano@est.iue.edu.ar, editxgarcia@gmail.com,
maiorano@gmail.com, edumalvacio@gmail.com, tapianestor87@gmail.com
nico_n44@hotmail.com; pablo_1711_pap@hotmail.com

1. Resumen.

Esta línea de investigación busca la resolución de problemas abiertos acerca de la complejidad lineal y período de las secuencias binarias seudoaleatorias.

Las mismas pueden ser generadas, por ejemplo, por Registros de Desplazamientos Realimentados No Linealmente (NLFSR: Non-Linear Feedback Shift Registers). En particular los algoritmos Trivium [1, 2] y Trivium Toy[3].

El algoritmo Trivium ha conformado el portfolio final del concurso europeo e-Stream del año 2005 [4].

A la fecha, no se conocen ataques efectivos contra este generador [5, 6, 7].

El estudio de las propiedades de las secuencias seudoaleatorias en general puede extenderse a cualquier algoritmo, por ejemplo la familia A5 -empleados en telefonía celular-, los Generadores Controlados por Reloj (Clock-Controlled Generators) u otros.

El objetivo es lograr un estudio completo de los fundamentos matemáticos involucrados. Así poder medir la robustez criptológica de los generadores de secuencias seudoaleatorias.

La teoría de los campos finitos y los registros de Desplazamientos Lineales (LFSRs) y No Lineales (NLFSR) [8] ofrecen las herramientas matemáticas para abordar las problemáticas involucradas.

Palabras Claves:

Randon Sequences. Stream Ciphers. LFRS. NLFSR.

2. Contexto.

El CRIPTOLAB (Laboratorio de Investigación en Técnicas Criptográficas y Seguridad Teleinformática) pertenece a la Escuela Superior Técnica “Gral. Div. Manuel N. Savio” (EST), Facultad del Ejército, Universidad Nacional de la Defensa (UNDEF) en el área de la carrera de grado de Ingeniería en Informática y el posgrado en Criptografía y Seguridad Informática que se dictan en esta institución.

El desarrollo científico y tecnológico es relevante a nivel estratégico y es por ello que tanto las Fuerzas Armadas en general como el Ejército en particular destina recursos de investigación para cumplir con tal fin.

Resultados parciales de esta investigación han sido presentados en CACIC 2013 [3] y CACIC 2014[9], siendo ambas presen-

taciones premiadas como “mejor exposición” del Workshop de Seguridad Informática.

Asimismo dicho trabajo ha sido seleccionado entre los mejores del mencionado congreso y hemos sido invitados a incluirlo en el número regular del Journal of Computer Science and Technology [10].

Además hemos podido dar entidad propia a esta línea de investigación al poder incluirla dentro de los proyectos de la EST – IUE bajo el nombre de Stream Cipher: Estudio de las propiedades y vulnerabilidades de generadores pseudoaleatorios de la familia Trivium.

3. Introducción.

Los generadores lineales del tipo Linear Feedback Shift Register (LFSR) generan secuencias pseudo-aleatorias con período y complejidad lineal controladas. Su estudio comenzó alrededor de los años '60 [11, 12].

Dada su naturaleza lineal, los LFSRs resultan ser inseguros: cuando $2n$ bits consecutivos (siendo n la longitud del registro) de la secuencia de salida de un LFSR son conocidos, toda la sucesión resulta ser predecible. Sistemas basados en LFSRs intentan agregar complejidad linealidad combinando, entre otras cosas, sus salidas a través de una función no lineal. Esto tampoco ofrece la seguridad deseada.

Los NLFSRs (una generalización de los LFSRs) estuvieron por mucho tiempo postergados en la comunidad criptológica. Su estudio se revitalizó con el advenimiento de la llamada “Criptografía Liviana”. La misma puede ejecutarse sobre plataformas de poco poder de cálculo. Incluso una cantidad de nuevos dispositivos tales como marcapasos, procesadores centrales montados en autos, grúas, tractores

y cosechadoras de alto desempeño, entre otros.

La revitalización del estudio de los NLFSRs ha comenzado a aparecer literatura. Tal es el caso de la familia TRIVIUM (De Cannière-Preneel), BIVIUM, CUADRIVIUM, entre otros.

4. Líneas de Investigación, Desarrollo e Innovación

Inicialmente el estudio del algoritmo Trivium y el Trivium Toy [13] llevó los investigadores a desarrollar herramientas propias e implementar otras reconocidas por la comunidad científica, desde los fundacionales test de Golomb hasta las baterías de test del NIST, “DieHard”¹ y “DieHarder”².

Estas herramientas permiten abordar los problemas de la complejidad lineal, la búsqueda del período y demás propiedades criptológicas que se desean conocer de una determinada secuencia pseudoaleatoria.

El CriptoLab podría analizar las propiedades criptológicas de secuencias de números generado por un modelo basado en álgebras no conmutativas. En particular el esquema que utiliza como elementos de la estructura algebraica a los Cuaterniones de Hamilton [14].

5. Resultados y Objetivos.

Se espera poder analizar y evaluar las características y propiedades criptológicas de enormes secuencias de números.

6. Formación de Recursos Humanos.

Además de los investigadores que forman parte del staff fijo del laboratorio, el

¹ <http://stat.fsu.edu/pub/diehard/>

² <http://www.phy.duke.edu/~rgb/General/dieharder.php>

equipo de investigación cuenta con la participación de 2 estudiantes del posgrado en Criptografía y Seguridad Teleinformática. Los mismos están realizando sus Trabajos Finales de Integración (tesina de posgrado) en temas afines a esta línea de investigación, en la cual colaboran.

Durante el año 2015 se han sumado al proyecto dos alumnos de 4to año de la carrera de Ingeniería Informática. Uno de ellos es Pablo Pérez como alumno ayudante y Nicolás Dulio que ha recibido la beca Estímulo a las Vocaciones Científicas, perteneciente al Programa Estratégico de Investigación y Desarrollo, Plan de Fortalecimiento (Componente de Formación de Recursos Humanos) del Consejo Interuniversitario Nacional.

7. Referencias y Bibliografía

- [1] De Cannière, C. and Preneel, B. “TRIVIUM A Stream Cipher Construction Inspired by Block Cipher Design Principles”. In Workshop on Stream Ciphers Revisited (SASC2006), 2006.
- [2] De Cannière, C. and Preneel, B. “TRIVIUM Specifications”. eSTREAM, ECRYPT Stream Cipher Project, Report. 2008.
- [3] Castro Lechtaler, A.; Cipriano, M.; García, E.; Liporace, J.; Maiorano, A.; Malvacio, E. “Model Design for a Reduced Variant of a Trivium Type Stream Cipher”. XIX Congreso Argentino de Ciencias de la Computación, Mar del Plata, Buenos Aires. 2013.
- [4] eSTREAM: eSTREAM – The ECRYPT Stream Cipher Project: <http://www.ecrypt.eu.org/stream/>
- [5] McDonald, C. and Pieprzyk, C. “Attacking Bivium with MiniSat”, Cryptology ePrint Archive, Report 2007/040, 2007.
- [6] Raddum, H. “Cryptanalytic Results on Trivium”, eSTREAM, ECRYPT Stream Cipher Project, Report 2006/039, 2006.
- [7] Maximov, A. and Biryukov, A. “Two Trivial Attacks on Trivium”, Selected Areas in Cryptography, Lecture Notes in Computer Science, Vol.4876, Springer, 2007.
- [8] Dubrova, E. “A List of Maximum-Period NLFSRs”, Cryptology ePrint Archive, Report 2012/166, March 2012, <http://eprint.iacr.org/2012/166>
- [9] Castro Lechtaler, A.; Cipriano, M.; García, E.; Liporace, J.; Maiorano, A.; Malvacio, E., Tapia, N., “Trivium Vs. Trivium Toy”. XX Congreso Argentino de Ciencias de la Computación, Octubre 2014. Universidad Nacional de La Matanza, San Justo, Buenos Aires.
- [10] Castro Lechtaler, Antonio; Cipriano, Marcelo; García, Edith; Liporace, Julio; Maiorano, Ariel; Malvacio, Eduardo. “Model design for a reduced variant of a Trivium Type Stream Cipher.” Journal of Computer Science and Technology Vol. 14, No. 1, Abril 2014.
- [11] Golomb. “Shift Register Sequences”. Aegean Park Press, 1982.
- [12] Massey, J.L. “Shift-register synthesis and BCH decoding”. IEEE Transactions on Information Theory 15, 1969.
- [13] Castro Lechtaler, A.; Cipriano, M.; García, E.; Liporace, J.; Maiorano, A.; Malvacio, E., Tapia, N., “On the Interleaving Process Applied to the Trivium Algorithm”. II Congreso Nacional de Ingeniería Informática/Ingeniería de Sistemas (CoNaIISI), Noviembre 2014. Universidad Nacional de San Luis, San Luis.
- [14] Kamlofsky, J; Hecht, J; Hidalgo Izzi, O; Abdel Masih, S. “A Diffie-Hellman Compact Model Over Non-Commutative Rings Using Quaternions”, VIII CIBSI-TIBETS, Ecuador, 2015.

Optimización de las fórmulas para la detección de Infraestructuras de Clave Pública anómalas.

Castro Lechtaler, Antonio^{1,2}; Cipriano, Marcelo¹; Malvacio, Eduardo¹;
Dulio, Nicolás¹; Pérez, Pablo¹

¹Laboratorio de Investigación en Técnicas Criptográficas y Seguridad Teleinformática.
Escuela Superior Técnica, Facultad de Ingeniería. Instituto Universitario del Ejército.

²CISTIC/FCE - Universidad de Buenos Aires.

acastro@est.iue.edu.ar, marcelocipriano@est.iue.edu.ar, edumalvacio@gmail.com
nico_n44@hotmail.com

1. Resumen.

Este proyecto de investigación persigue elaborar un software de auditoría que permite detectar comportamientos anómalos en una Infraestructura de Clave Pública (Public Key Infrastructure: PKI por sus siglas en inglés) en lo concerniente a la calidad de los números primos en los certificados emitidos por la entidad.

Dada la imposibilidad de abarcar todo el universo de certificados posibles que una PKI puede emitir, se llevará adelante un análisis estadístico de un conjunto de muestras formadas por certificados entregados, basado en las propiedades matemáticas de la densidad y distribución de números primos de determinado tamaño.

La existencia de sesgos o problemas en la generación y/o selección de los números primos que constituyen a cada certificado, puede devenir en vulnerabilidades susceptibles de ser explotadas. Afectando de esta manera el control de acceso al sistema, el intercambio de claves para sesiones seguras, problemas con la autenticación de usuarios, mensajes y equipos, etc.

El estudio y análisis de las propiedades matemáticas involucradas y la determinación de criterios adecuados a la detección

del comportamiento anormal permitirán desarrollar un software auditor de PKI.

Al día de hoy no existen fórmulas que generen números primos ni se conoce con exactitud la forma en que dichos números se distribuyen. Asimismo la cantidad de dichos números y el tamaño de los mismos agregan un componente complejo al problema. Y ello sin mencionar la dificultad computacional para realizar cálculos precisos con ellos.¹

Palabras Clave:

Seguridad en Redes, Infraestructura de Clave Pública, PKI, Detección de Anomalías, Open-SSL, RSA.

2. Contexto.

El CriptoLab (Laboratorio de Investigación en Técnicas Criptográficas y Seguridad Teleinformática) pertenece a la Escuela Superior Técnica “Gral. Div. Manuel N. Savio” (EST), Facultad del Ejército, Universidad Nacional de la Defensa (UNDEF) en el área de la carrera de grado de Ingeniería en Informática y el posgrado en Criptografía y Seguridad Informática que se dictan en esta institución.

¹Calcular $n!$ con n del orden de 2048 bits, por ejemplo.

Las Fuerzas Armadas en general como el Ejército en particular destinan recursos a la investigación y el desarrollo científico-tecnológico por ser considerados relevantes a nivel estratégico.

En particular, el Instituto de Investigaciones Científicas y Técnicas para la Defensa (CITEDEF) realizó aportes a través de Proyectos de Investigación Científico Tecnológicos Orientados (PICTO) para la realización durante 6 años del proyecto recientemente finalizado sobre “REDES PRIVADAS COMUNITARIAS”.

En el marco de dicho proyecto se pudo realizar parte de esta investigación. Los resultados parciales o finales han sido presentados en varios CACIC (Congreso Argentino de Ciencias de la Computación) para su difusión y consideración de la comunidad científica.

Dentro de la EST esta línea de investigación se lleva adelante bajo el nombre de VULCLAP: Vulnerabilidades en Clave Pública.

CITEDEF y el Consejo Profesional de Ingeniería de Telecomunicaciones, Electrónica y Computación (COPITEC) dieron su aval por escrito para este proyecto, dado su interés en ser aplicado en sus propios sistemas y redes.

3. Introducción.

Los sitios de internet pueden autenticarse por medio de certificados digitales. De manera parecida los sistemas pueden componer una estructura que permita ese servicio y otros. Tal subsistema recibe el nombre de Infraestructura de Clave Pública (PKI).

Existen PKI de código abierto y libre e incluso algunas tienen licencia de uso gratuita.

¿Se puede comprobar si estos procesos y servicios contienen errores que pueden minar la seguridad? [1].

Una manera de hacerlo es acceder a su código fuente y revisar todas las líneas de programación y analizar su comportamiento. Está clara la complejidad de tal tarea y la dificultad (o imposibilidad, al menos hasta ahora) de automatizarlo.

Otra manera es el software que se lograría con el progreso de esta línea de investigación.

La filosofía del código abierto (open source) y la ley de Linux [2] por sí mismas no garantizan la ausencia de errores. Un ejemplo es el bug descubierto por Luciano Bello en OpenSSL2 de Debian. Pasaron 20 meses desde que la versión defectuosa fuera informada [3] hasta que se publicó un parche. Y este no es el único ejemplo. Recientemente han sido detectados problemas de seguridad en SSL/TLS en el iOS7 de Apple y GnuTSL de Linux, entre otros.³

Si una PKI generase un conjunto sesgado de números primos para emitir sus certificados, entonces un atacante podría vulnerar la seguridad de dicho sistema.

Lenstra y otros [4], han hallado repetición de los números primos en el 5% de una gran muestra de certificados digitales de 1024 bits. Si se repiten los primos, entonces tales certificados son vulnerables.

4. Líneas de Investigación, Desarrollo e Innovación.

Se pudo determinar una fórmula, con la que se elaboró una herramienta informática, que permite hallar los primos que

² Una mala inicialización de una variable provocó una predictibilidad en el generador de números, abriendo una vulnerabilidad inimaginable.

³ <http://www.securitynull.net/>

componen un módulo RSA con la información que aportan su clave pública y privada [5].

Las posteriores pruebas de codificación e implementación demostraron que este procedimiento corría muy veloz [6].

Se comparó el rendimiento del mismo con el procedimiento existente en la bibliografía tradicional para la enseñanza de la criptografía [7].

Los análisis indicaron que la complejidad computacional del algoritmo era del orden $O(\log n)$ mientras que el de la bibliografía de referencia tenía un orden $O(\log^3 n)$ [8].

Se elaboró la herramienta matemática que permitiría detectar anomalías [9].

Se realizó un abordaje probabilístico del problema [10] y por último el diseño final de la herramienta probabilística y estadística [11-13]

Simultáneamente al avance matemático se ensamblaron y codificaron todas las herramientas matemáticas antes mencionadas, en una plataforma de software programado en C++.

5. Resultados y Objetivos.

Se llevaron adelante pruebas por muestras o lotes, evaluándose 80.000.000 de módulos agrupados en 80000 lotes de 1000 cada uno.

Actualmente se está tratando de mejorar la fórmula matemática del modelo estadístico pues se necesita computar enormes factoriales y su resolución es, hasta ahora, muy costosa computacionalmente hablando.

Concluido este paso, se simularán diferentes PKI's con vulnerabilidades y sin ellas. Cada una será testeada por la herramienta de auditoría en elaboración. El comportamiento esperado es que de-

tecte las vulnerables e informe al respecto.

Se elaboró una plataforma de computación distribuida a los efectos de acelerar la investigación.

6. Formación de Recursos Humanos.

Desde el año 2012 algunos algoritmos que utilizamos en esta investigación fueron codificados y probados en el contexto de la Cátedra de Computación I a cargo del Ing. Mg. Alejandro Repetto, que posee nuestra facultad en la carrera de Ingeniería Informática.

Desde el año 2013 un equipo de docentes y alumnos del Centro de Investigación y Desarrollo de Software del Ejército Argentino (CIDESO) trabaja en el diseño y elaboración de una plataforma de computación distribuida para que se pueda emplear en problemas de criptología a los que se dedica el CriptoLab y se han publicado sus resultados [12, 14].

Durante el año 2015 se han sumado al proyecto dos alumnos de 4to año de la carrera de Ingeniería Informática. Uno de ellos es Pablo Pérez como alumno ayudante y Nicolás Dulio que ha recibido la beca Estímulo a las Vocaciones Científicas, perteneciente al Programa Estratégico de Investigación y Desarrollo, Plan de Fortalecimiento (Componente de Formación de Recursos Humanos) del Consejo Interuniversitario Nacional.

7. Referencias

[1] Young A and Yung M. *An Elliptic Curve Asymmetric Backdoor in Open-SSL RSA Key Generation*. Chapter 10. *Cryptovirology*. 2006.

<http://www.cryptovirology.com>.

- [2] Glass, Robert “*Facts and Fallacies of Software Engineering*”. Addison-Wesley Professional, 2003.
- [3] Bello L, Bertacchini M. “*Generador de Números Pseudo-Aleatorios Predecible en Debian*”. III Encuentro Internacional de Seguridad Informática. Manizales, Colombia. Octubre 2009.
- [4] Lenstra, A; Hughes, J; Augier, M y otros. Ron was wrong, Whit is right. e-print International Association for Cryptologic Research. 15 Feb 2012
<http://eprint.iacr.org/2012/064>,
- [5] Cipriano, M. “Factorización de N: recuperación de factores primos a partir de las claves pública y privada.” XIV Congreso Argentino de Ciencias de la Computación CACIC 2008. Chilecito, La Rioja, Octubre 2008.
- [6] Castro Lechtaler, C; Cipriano, M; Benaben A; Quiroga, P. “*Study on the effectiveness and efficiency of an algorithm to factorize N given e and d*”. IX Seminario Iberoamericano en Seguridad de las Tecnologías de la Información, La Habana, CUBA. 2009.
- [7] Menezes, A; van Oorschot, P and Vanstone, S. *Handbook of Applied Cryptography*. CRC Press. 5th Edition, 2001.
- [8] Benaben, A; Castro Lechtaler, A; Cipriano, M; Foti, A. “*Development, testing and performance evaluation of factoring algorithms whit additional information*” XXVIII Conferencia Internacional de la Sociedad Chilena de Computación. Santiago de Chile. 2009.
- [9] Castro Lechtaler, A; Cipriano, M. “*Detección de anomalías en Oráculos tipo OpenSSL por medio del análisis de probabilidades*”. XVII Congreso Argentino de Ciencias de la Computación CACIC 2011. La Plata, Buenos Aires, Octubre 2011.
- [10] Castro Lechtaler, Antonio, Cipriano Marcelo; Malvacio Eduardo; Cañón, Sebastián; *Procedure for the Detection of Anomalies in Public Key Infrastructure (RSA Systems)*. XIII Simposio Argentino de Tecnología, 41 Jornadas Argentinas de Informática e Investigación Operativa JAIIO – SADIO. La Plata, Buenos Aires, Agosto 2012.
- [11] Castro Lechtaler, Antonio; Cipriano, Marcelo; Malvacio, Eduardo. *Experimental detection of anomalies in public key infrastructure*. XVIII Congreso Argentino de Ciencias de la Computación CACIC 2012. Bahía Blanca, Buenos Aires, Octubre 2011.
- [12] Castro Lechtaler, A; Repetto, A; Bianchi, O; Cipriano, M; Arroyo Arzubi, A; Cicerchia, C; Malvacio, E. *ULTRACOM: Computación de alto rendimiento para criptoanálisis*. XX Congreso Argentino de Ciencias de la Computación, La Matanza, Buenos Aires, 2014.
- [13] Castro Lechtaler, Antonio; Cipriano, Marcelo; Malvacio, Eduardo. *Probabilidades de referencia para aplicar en la detección de Infraestructuras de Clave Pública anómalas* XXI Congreso Argentino de Ciencias de la Computación, Junín, Buenos Aires, 2015.
- [14] Castro Lechtaler, A; Repetto, A; Bianchi, O; Cipriano, M; Arroyo Arzubi, A; Cicerchia, C; Malvacio, E. *Computación en grilla de escritorio para evaluación de algoritmos criptográficos*. XVII Workshop de Investigadores en Ciencias de la Computación, Salta, 2015.

AVANCES EN EDUCACIÓN DE DINÁMICA DE TECLEO Y EL CONTEXTO EMOCIONAL DE UN INDIVIDUO APLICANDO INTERFAZ CEREBRO COMPUTADORA

Enrique P. Calot¹, Federico M. Rossi¹, Ariel Liguori¹, Nahuel González¹, Waldo Hasperu², Jorge S. Ierache¹

² Instituto de Investigación en Informática LIDI
Facultad de Informática - Universidad Nacional de La Plata

¹Laboratorio de Sistemas de Información Avanzados,
Departamento de Computación, Facultad de Ingeniería,
Universidad de Buenos Aires

Av. Paseo Colón 850 - C1063ACV - Buenos Aires - Argentina
Tel +54 (11) 4343-0893 / 4343-0092.
{ecalot, jierache}@lsia.fi.uba.ar

Resumen

El presente artículo describe avances y objetivos del proyecto durante el último año. Orientado a la articulación entre el patrón biométrico de cadencia de tecleo y la determinación de estados emocionales, el proyecto se enfocará en desarrollar un marco de trabajo que permita determinar personas y las relaciones de las mismas con su estado afectivo. Para el mismo se está trabajando en obtener un banco de datos adquirido con una interfaz cerebro-computadora sobre un individuo al teclado. Este entorno de trabajo permitirá analizar las modelizaciones de las emociones adquiridas y plantear los cambios temporales de ritmos de tecleo de un usuario y analizarlos en función de sus emociones. En este contexto fue desarrollado un prometedor algoritmo de análisis de tecleo en textos libres que permite identificar personas, un robusto set de datos de dinámicas de tecleo y un ambiente de experimentación con el que se recabarán los datos emocionales y dinámicas de tecleo.

Palabras clave: *Dispositivos de bioseñales, Educación de emociones, Dinámica de tecleo, Contextos finitos.*

Contexto

El Proyecto articula líneas de trabajo de Keystroke Dynamics y Brain Machine Interface (BMI) orientados tanto a la seguridad como a la extracción de patrones emocionales en el ser

humano. El proyecto se desarrolla en el marco del UBACyT 2014–2017 GEF, con radicación en el Laboratorio de Sistemas de Información Avanzados de la Facultad de Ingeniería de la Universidad de Buenos Aires. El mencionado proyecto se realiza con la colaboración del ISIER de FICCTE UM, apoyando específicamente en el área de aplicación de BMI, en el marco del PID 01-001/14. El proyecto vincula un tesista de doctorado de UNLP.

Introducción

La firma manuscrita tiene su paralelo en el teclado: los patrones neurofisiológicos que vuelven única a una firma manuscrita se pueden observar también en el ritmo de tecleo de un usuario [Joyce & Gupta, 1990; Calot et al., 2015]. La técnica que analiza este tipo de patrones se llama Dinámica de Tecleo o Keystroke Dynamics. Se ha realizado una comparación de algoritmos que utilizan dicha técnica [Killourhy & Maxion, 2009] para detectar usuarios e impostores. También existen estudios que vinculan los ritmos de tipeo con el estado emocional de las personas [Epp et al., 2011; Calot et al., 2015].

A través del modelado de dinámicas de tecleo y la utilización de clasificadores de última generación se han alcanzado resultados excelentes en la clasificación de claves fijas. No es el caso en el modelado de textos libres; si bien existen avances promisorios [Gunetti & Picardi, 2005] no se ha logrado una similar eficacia en la clasificación ni se conoce si es factible lograrlo, considerando por

ejemplo la dificultad de la variación del idioma de escritura.

Una Interfaz Cerebro—Computadora (Brain—Machine Interface, BMI) facilita la comunicación entre las funciones mentales o cognitivas creadas a partir del cerebro de una persona, captando las señales eléctricas, para ser procesadas, clasificadas y comunicadas con aplicaciones o dispositivos específicos [Ierache et al. 2012, 2013a, 2013b; 2014]. En el contexto de emociones, el trabajo de [Ierache et al., 2015] presenta los primeros resultados del estado emocional centrado en estímulos, aplicando interfaz cerebro—maquina.

En Dinámica de Tecleo se obtuvo un set de datos de prueba y un prometedor algoritmo de clasificación de patrones en textos libres [González, 2016]. Además se ha trabajado en la articulación de ambos temas diseñando un ambiente de captura de datos que servirá para la creación de un nuevo conjunto de datos de prueba que contendrá patrones de tecleo y estados emocionales obtenidos mediante tecnología BMI.

Líneas de Investigación, Desarrollo e Innovación

En el actual proyecto se desarrollan varias líneas de investigación, varias relacionadas entre sí, las cuales se detallan a continuación.

Experimentar con interfaz cerebro—maquina—captura de emociones

Con el fin de complementar el desarrollo de las distintas líneas de investigación mencionadas se está experimentando con mecanismos y modelos que permiten la lectura de la actividad cerebral y el procesamiento de datos crudos para obtener información acerca de los estados emocionales de los usuarios.

Experimentar con ritmos de tecleo

En esta línea de investigación se estudian algoritmos para clasificar e identificar personas mediante sus patrones de tecleo.

Desarrollar un ambiente de experimentación real

Para modelos que permitan estudiar la existencia de patrones entre los estados emocionales y la dinámica de tecleo de una persona, es de suma importancia establecer un marco experimental en el cual sea posible construir un conjunto de datos que goce de niveles de aleatoriedad reducidos, imponiendo en los usuarios situaciones medibles que reduzcan la brecha que

divide a las emociones y al comportamiento exteriorizado del tipoo.

Efectuar la experimentación, crear un entrenamiento y un banco de datos con información acerca del ritmo de tipoo según estados emocionales y analizar sus resultados

Una vez armado el ambiente se procederá a hacer una experimentación con voluntarios, determinar relaciones entre el estado emocional de un individuo adquirido por su dinámica de tecleo y los datos adquiridos con una interfaz cerebro—computadora (BMI).

Resultados y Objetivos

A continuación se detallan objetivos y resultados de las líneas descritas en el apartado anterior.

Experimentar con interfaz cerebro—maquina—captura de emociones

La lectura y adquisición de datos relacionados a la actividad cerebral ha sido posible gracias a la utilización del Emotiv EPOC, un dispositivo conformado por electrodos dispuestos en el sistema internacional 10-20 [Ierache et al., 2013^b], capaz de transmitir las señales detectadas a una interfaz controlada por computadora.

Los datos son recogidos por OpenViBE, un lenguaje de programación gráfico de código abierto que está diseñado principalmente para desarrollar interfaces cerebro—máquina. Además de su potencialidad, una gran ventaja es que permite a los usuarios trabajar con más de quince dispositivos de adquisición de EEG, entre ellos, el Emotiv EPOC [Yurci, 2014].

En nuestros experimentos, OpenViBE es utilizado para, a partir de las señales recibidas desde el dispositivo, aplicar filtros pasabanda que nos permitan obtener las bandas de frecuencias *alpha* y *beta* de cada uno de los electrodos. Estos valores pueden ser utilizados para calcular la posición del estado emocional en el modelo bidimensional de excitación—valencia [Feldman, 1995; Oude Bos, 2006] (figura 1).

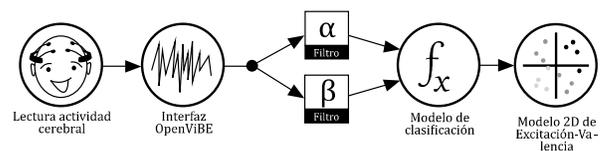


Figura 1. Adquisición de actividad cerebral y clasificación de emociones.

Dado que los objetivos no se centran en desarrollar nuevos modelos de detección de estados emocionales a partir de EEG, es que se está ahondando en distintos trabajos de investigación relacionados a dicha temática. El fin de esto es tomar conocimiento de aquellos modelos que han demostrado ser más exactos en la clasificación de emociones y, de esta manera, utilizarlos para la conformación del conjunto de datos que relaciona la dinámica de tecleo y los estados emocionales.

Actualmente se encuentra presente la problemática de contar con varios modelos que utilizan las bandas *alpha* y *beta* para mapear dentro del modelo de excitación-valencia [Ramirez & Vamvakousis, 2012; Giraldo & Ramirez, 2013; Yurci, 2014], pero para todos ellos aún no hemos encontrado contrastado sus experimentos para confirmar el sustento teórico que otorguen confiabilidad al momento de ser aplicados. Por esta razón, nos encontramos en proceso de búsqueda de dichas validaciones teóricas.

Experimentar con ritmos de tecleo.

Se realizaron publicaciones en distintos congresos internacionales. En [Calot et al., 2014] se había mostrado que, si bien en resultados previos en comparación de algoritmos para calcular la distancia entre patrones de tecleo en textos fijos la distancia de Mahalanobis resultó muy prometedora, en realidad lo era porque el conjunto de datos utilizado para homogeneizar la prueba era linealmente dependiente entre sí y la distancia de Mahalanobis eliminaba la columna dependiente; esto favorece al algoritmo más simple (la distancia euclidiana normalizada) que tuvo una performance incluso un 0,24% superior [Calot et al., 2015].

En [González & Calot, 2015] fueron presentados resultados de un nuevo algoritmo propuesto en el entorno de esta investigación para análisis de patrones de cadencia de tecleo en textos libres, donde se lograron resultados superadores. En este contexto se desarrolló la tesis “Utilización de contextos finitos para el modelado de cadencias de tecleo en esquemas de autenticación mixta” [González, 2016] defendida el 29 de febrero de 2016.

Adicionalmente, fue enviado para su evaluación el artículo *A Replication of Two Free Text Keystroke Dynamics Experiments under Harsher Conditions* a un congreso internacional. Aquí se realizó una réplica de los datos obtenidos en los trabajos de [Gunetti & Picardi, 2005; Bergadano et al., 2002] en un contexto mucho más real con una base de datos producida en el contexto de esta investigación [Calot et al., 2015] y luego los

resultados fueron comparados con los obtenidos por el algoritmo de [González, 2016]. Los resultados para el algoritmo desarrollado fueron prometedores.

Desarrollar un ambiente de experimentación real

Se ha construido una aplicación web capaz de leer la dinámica de tecleo y la actividad cerebral a medida que el usuario interactúa con el entorno presentado, en el cual se lo estimula visualmente a fin de inducirlo y predisponerlo a ciertos estados emocionales.

De esta manera, el objetivo principal del experimento se centra en crear un conjunto de datos realista que dé paso al desarrollo de modelos que mejoren los actuales porcentajes de acierto a la hora de identificar los estados emocionales de una persona a partir de su dinámica de tecleo.

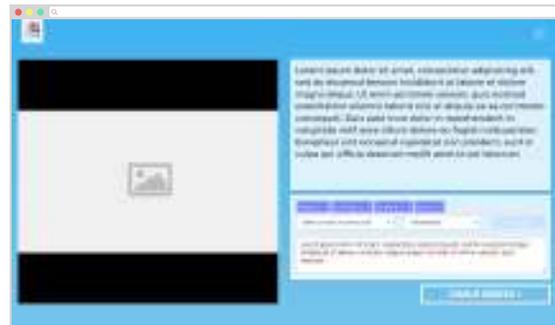


Figura 2. Interfaz principal de la aplicación utilizada en el experimento.

Más específicamente, el ambiente de experimentación (figura 2) consta de un visualizador multimedial (columna izquierda de la figura) en el que se proyectan imágenes o videos para inducir estados emocionales; un área de texto donde se presentan fragmentos varios tomados una selección propia (intentando garantizar la generación de una emoción independientemente del entorno cultural e idioma nativo de la persona); un campo de texto donde los usuarios deberán transcribir dichos fragmentos y un selector de emoción-intensidad optativo según el experimento que permitiría a las personas expresar su propia percepción de las emociones que se están experimentando.

La captura de datos está compuesta por un conjunto de fases experimentales en las cuales se inducen uno a uno distintos estados emocionales, basado en el modelo emocional 2D de excitación-valencia. La tecnología BMI a utilizar será EPOC de Emotiv y para su posterior análisis se utilizará OpenViBE; ambos mencionados en la sección de

experimentación con BMI. En cada fase se expondrán imágenes provistas por la librería IAPS [Lang, et al. 2008] y se mostrará un fragmento de texto cuyo contenido tiene como propósito potenciar la inducción emocional de las imágenes. De esta manera, se logra acaparar la atención de los usuarios mientras transcriben el texto.

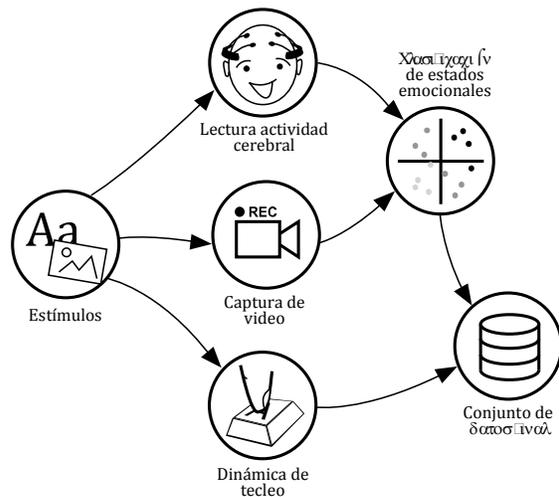


Figura 3. Ciclo simplificado de recolección de datos.

A medida que transcurren cada una de las etapas, la dinámica de tecleo, la lectura de la actividad cerebral y la captura de las expresiones faciales del usuario son leídas y almacenadas (figura 3).

Para analizar gestos emocionales en el video del usuario se utilizará la herramienta *clmtrackr* [Øygard, 2015], que detecta caras y luego interpreta emociones a partir de un set de datos de puntos mapeados de una base de datos. Su implementación utiliza modelos locales restringidos ajustados por puntos de referencia regularizados (*constrained local models fitted by regularized landmark mean-shift*) desarrollado por Audun Mathias Øygard y basado en el artículo [Saragih et al., 2011]. Para detectar los estados emocionales se utilizó la base de datos MUCT [Milborrow et al., 2010] agregando algunas imágenes propias de Øygard.

Efectuar la experimentación, crear un entrenamiento y un banco de datos con información acerca del ritmo de tipeo según estados emocionales y analizar sus resultados

Para realizar la experimentación hemos solicitado un conjunto de datos recopilado por el *Center for the Study of Emotion & Attention* de la *University of Florida* en *Gainesville, FL, Estados*

Unidos. La solicitud aún se encuentra pendiente. El conjunto contiene una base de datos de imágenes (IAPS [Lang, et al. 2008]) y sonidos (IADS [Bradley & Lang, 1999]) cuya distribución debe ser solicitada a esta entidad, sus datos no son de distribución pública para evitar el reconocimiento por parte de los usuarios de las imágenes y sonidos en caso de haber visto las imágenes previamente. Debido a esta medida la base de datos es muy utilizada en estudios científicos que analizan emociones. Una vez recibidas las imágenes y los audios procederemos a utilizar el ambiente para la captura del set de datos con personas.

Una vez capturado este banco de datos nuestro objetivo será determinar relaciones entre el estado emocional de un individuo adquirido por su dinámica de tecleo y los datos adquiridos con una interfaz cerebro-computadora (BMI) y los correlacionados por las imágenes de video de su rostro. Esperamos comenzar a publicar resultados una vez terminado el análisis de resultados del set de datos producido en las experimentaciones.

Formación de Recursos Humanos

El equipo de trabajo se conforma de dos investigadores formados, dos investigadores en formación y dos investigadores alumnos de grado. En el marco de las líneas de investigación se: desarrollo una tesis de grado presentada este año en la Facultad de Ingeniería de la UBA; se cuenta con un tesista doctoral en la UNLP radicado en LSIA de la UBA y dos investigadores por presentar sus proyectos de tesis doctoral en UNLP y de grado en FIUBA. Adicionalmente, se ha propiciado la participación de alumnos de grado para la realización de otros trabajos dentro de las líneas de investigación.

Referencias

- BERGADANO, F.; GUNETTI, D.; & PICARDI, C. (2002). User Authentication Through Keystroke Dynamics. *ACM Transactions on Information and System Security (TISSEC)*, 5(4), 367–397. <http://doi.acm.org/10.1145/581271.581272>
- BRADLEY, M. M.; & LANG, P. J. (1999). International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Tech. Rep. No. B-2). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.
- CALOT, E. P.; ACETO, E. L.; RODRÍGUEZ, J. M.; LIGUORI, A.; OCHOA, M. A.; MERLINO, H.; FERNÁNDEZ, E.; GONZÁLEZ, N.; PIRRA, F.; & IERACHE, J. S. (2015). Líneas de investigación del Laboratorio de Sistemas de Información Avanzados. In *Proceedings del XVII*

- Workshop de Investigadores en Ciencias de la Computación (pp. 4 p.).
- CALOT, E. P.; RODRÍGUEZ, J. M.; & IERACHE, J. S. (2014). Improving versatility in keystroke dynamic systems. In *Computer Science & Technology Series. XIX Argentine Congress of Computer Science, Selected papers*. ISBN 978-987-1985-49-4, Editorial de la Universidad Nacional de La Plata (EDULP).
- EPP, C.; LIPPOLD, M.; & MANDRYK, R. L. (2011). Identifying emotional states using keystroke dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2011* (pp. 715–724). ACM.
- FELDMAN, L. A. (1995). Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of personality and social psychology*, 69(1), 153.
- GIRALDO, S.; & RAMIREZ, R. (2013). Brain-activity-driven real-time music emotive control. In *Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11th-15th June 2013*. Geoff Luck & Olivier Brabant (Eds.). ISBN 978-951-39-5250-1. University of Jyväskylä, Department of Music.
- GONZÁLEZ, N.; & CALOT, E. P. (2015). Finite Context Modeling of Keystroke Dynamics in Free Text. In *Biometrics Special Interest Group (BIOSIG), 2015 IEEE International Conference of the* (pp. 1–5).
- GONZÁLEZ, N. (2016). Utilización de contextos finitos para el modelado de dinámicas de tecleo en esquemas de autenticación mixta. Tesis de Ingeniería en Informática, Facultad de Ingeniería, UBA.
- GUNETTI, D.; & PICARDI, C. (2005). Keystroke Analysis of Free Text. *ACM Transactions on Information and System Security (TISSEC)*, 8(3), 312–347. <http://doi.acm.org/10.1145/1085126.1085129>
- IERACHE, J.; PEREIRA, G.; IRIBARREN, J.; SATTOLO, I. (2013^a). Robot Control on the Basis of Bio-electrical Signals. *International Conference on Robot Intelligence Technology and Applications (RiTA), 2012*. Gwangju, Korea on December 16–18, 2012. *Series Advances in Intelligent and Soft Computing of Springer*.
- IERACHE, J.; PEREIRA, G.; CALOT, E.; IRIBARREN, J. (2013^b). Framework for Brain Computer Interface implemented to control devices in the context of home automation. *XIX Congreso Argentino de Ciencias de la Computación*, ISBN 978-897-23963-1-2.
- IERACHE, J.; PEREIRA, G.; IRIBARREN, J. (2012). Demostración de los resultados en la integración de Interfases Lectoras de Bioseñales aplicadas al Control de un Robot. *VII Congreso Educación en Tecnología y Tecnología en Educación Universidad Nacional del Noroeste de la Provincia de Buenos Aires*. UNNOBA, demos educativas. ISBN 978-987-28186-3-0.
- IERACHE, J.; PEREIRA, G.; IRIBARREN, J. (2014). Navigation Control of a Robot from a remote location via the Internet using Brain-Machine Interface; *Robot Intelligence Technology and Applications 2*. : Springer. 2014. p -297-310. ISBN 978-3-319-05581
- IERACHE, J.; NERVO, F.; PEREIRA, G.; IRIBARREN, J. (2015). Emotional Status Focused on Stimuli by Applying Brain-Machine Interface ,*Computer Science and Technology Series Selected Papers XX Congreso Argentino de Ciencias de la Computación*, pp 245-256, Edulp . Edition 2015 , ISBN 978-987-1985-71-5
- JOYCE, R.; & GUPTA, G. (1990). Identity authentication based on keystroke latencies. *Commun. ACM*, 33(2), 168–176, febrero 1990. <http://doi.acm.org/10.1145/75577.75582>
- KILLOURHY, K. S.; & MAXION, R. A. (2009). Comparing Anomaly-Detection Algorithms for Keystroke Dynamics. In *International Conference on Dependable Systems & Networks (DSN-09), Estoril, Lisbon, Portugal, 29 June–02 July, 2009* (pp. 125–134). IEEE Computer Society Press, Los Alamitos, California.
- LANG, P.J.; BRADLEY, M.M.; & CUTHBERT, B.N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.
- MILBORROW, S.; MORKEL, J.; & NICOLLS, F. (2010). The MUCT Landmarked Face Database. *Pattern Recognition Association of South Africa*.
- OUDE BOS, D. (2006). EEG-based emotion recognition- The Influence of Visual and Auditory Stimuli. *Capita Selecta (MSc course)*.
- ØYGARD, A. M. (2015). Emotion detection example. *clmtrackr*. URL: http://auduno.github.io/clmtrackr/examples/clm_e_motiondetection.html. Vigente al 2 de febrero de 2016.
- RAMIREZ, R.; & VAMVAKOUSIS, Z. (2012). Detecting emotion from EEG signals using the emotive eproc device. In *Brain Informatics* (pp. 175-184). Springer Berlin Heidelberg.
- SARAGIH, J. M.; LUCEY, S.; & COHN, J. F. (2011). Deformable Model Fitting by Regularized Landmark Mean-Shift. *Int. J. Comput. Vision*, 91(2), 200–215. <http://dx.doi.org/10.1007/s11263-010-0380-4>
- YURCI, E. (2014). Emotion Detection From EEG Signals: Correlating Cerebral Cortex Activity with Music Evoked Emotion. Tesis de Maestría. Universitat Pompeu Fabra, Barcelona.

Estructuras algebraicas aplicables en criptografía

Ignacio Gallego Sagastume

Facultad de Informática, Universidad Nacional de La Plata

ignaciogallego@gmail.com

Resumen

Existen diferentes estructuras algebraicas que son muy útiles en el ámbito de la seguridad informática y en particular en aplicaciones criptográficas. Algunas de estas estructuras son los cuadrados Latinos (CLs o LSs) y los quasi-grupos (QGs). Ellos son convenientes para la implementación de algoritmos de cifrado, de autenticación y protocolos de comunicación seguros.

Más específicamente, los CLs aleatorios sirven como claves de algoritmos de cifrado simétrico; además, son difíciles de generar y son inviables los ataques por fuerza bruta para adivinar un CL, dado el gigantesco espacio de claves que conforman (ver [MW05]).

El primer problema que se presenta es el de generar CLs aleatorios y uniformemente distribuidos, mediante algoritmos eficientes. Este es un problema parcialmente resuelto por [JM96] en el año 1996, pero existen más alternativas posibles según el grado de aleatoriedad deseado y la eficiencia computacional requerida en la generación. Se han implementado métodos propios con éxito en el proyecto [Gal15c]. La línea de investigación debe continuar con la exploración e implementación de diversos métodos ya conocidos para generación de estas estructuras, como pueden ser [Kos02],

[MW91] y [SVT14]. También se prevé trabajar sobre otras aplicaciones prácticas de estas estructuras en criptografía.

Palabras clave: generación, cuadrados Latinos, aleatorios, quasigrupos, QGs, LSs, CLs, distribución uniforme

Contexto

El proyecto en el que se enmarca esta investigación, comenzó a principios de 2012 con el trabajo de tesis de Magister [Gal15d], presentado en agosto de 2015 en la Facultad de Informática de la UNLP. El trabajo de tesis fue desarrollado bajo la dirección de la Dra. Claudia Pons (LIFIA, UNLP).

Desde aquél momento y hasta la actualidad, éste es un trabajo de investigación independiente y unipersonal, que carece de fuentes de financiamiento.

Introducción

Un cuadrado Latino (CL) de orden n es una matriz de dimensiones $n \times n$, que se completa utilizando n símbolos (por ejemplo los números de 0 a $n-1$), y en donde cada símbolo aparece exactamente una vez en cada fila y una vez en cada columna. Por ejemplo:

1 3 2
2 1 3
3 2 1

es un CL de orden 3. De un CL puede decirse si es aleatorio o no. Por ejemplo,

1 2 3 4
2 3 4 1
3 4 1 2
4 1 2 3

no es aleatorio, pues sus símbolos aparecen siempre en el mismo orden (informalmente no está “mezclado”).

Un algoritmo genera CLs aleatorios y uniformemente distribuidos si cualquier CL posible tiene la misma probabilidad de obtenerse como resultado (no hay tendencias en la generación). Esta propiedad la tiene el algoritmo de Jacobson & Matthews [JM96] (abreviado J&M), donde se demuestra que los resultados son *aproximadamente* uniformes.

La cantidad de todos los LSs posibles de orden n , llamada $L(n)$, es tan grande que en la actualidad solo se conoce hasta $L(11)$. A medida que n crece, $L(n)$ crece exponencialmente. Para órdenes mayores a 11 sólo se conocen cotas superiores e inferiores de $L(n)$ (ver [MW05]). Calcular este valor para cualquier n se trata de un problema de combinatoria sumamente complejo que en la actualidad permanece sin resolver. Esto implica que, para órdenes grandes (digamos, mayores a 16), un ataque por fuerza bruta probando todos los CLs resulta imposible (con el poder de cómputo disponible actualmente).

Los CLs pueden utilizarse como claves de algoritmos simétricos para cifrar y descifrar información en un protocolo criptográfico de comunicación. Los CLs aleatorios de orden 256 son de particular

importancia, porque permiten cifrar y descifrar cualquier carácter de la tabla de códigos ASCII, utilizando un recorrido secuencial a la manera de Gibson [Gib12].

Estas claves deben ser generadas con frecuencia para evitar posibles ataques. El primer problema que se presenta entonces es el de obtener un algoritmo que genere CLs aleatorios y aproximadamente uniformemente distribuidos de orden 256 y en una forma lo más eficiente posible.

Este problema fue parcialmente resuelto por J&M, pero, al implementar su algoritmo para órdenes tan grandes, el mismo resulta bastante ineficiente (se desempeña en $O(n^3)$).

Se han propuesto varias alternativas propias a este método en el trabajo de tesis [Gal15d], las cuales se describen en la sección de resultados. Algunas de éstas son más eficientes aunque menos uniformes. En la siguiente sección se detallan las líneas abiertas de investigación y objetivos a corto plazo, mientras que en la sección de resultados y objetivos se dan los objetivos a largo plazo.

Líneas de Investigación, Desarrollo e Innovación

Actualmente, se está trabajando en los algoritmos propuestos por McKay y Wormald [MW91] (pioneros en el área), que datan de antes del método propuesto por J&M y dan una implementación de tiempo de $O(nk^3)$, dado $k=o(n^{1/3})$, para generaciones de rectángulos Latinos de k filas por n columnas. Los resultados de este método son uniformemente distribuidos.

También se está trabajando en un método reciente propuesto por Selvi (et. al.) en [SVT14], que corrige un método

de O'Carroll de 1963 ([Car63]) que fallaba para algunos casos.

Otro enfoque al problema de generación es el de Fontana ([Fon13]), que extrae un CL uniformemente de todos los CLs posibles mediante un método todavía ineficiente en la práctica (sólo se ha llegado a generar CLs de orden 7 con este método). Pero aunque no sea práctico, los elementos expuestos en este trabajo resultan enriquecedores en el área y aportan elementos de análisis al problema de la generación.

La comprensión de estos trabajos en profundidad es esencial para su implementación en un lenguaje de programación, y para su posterior optimización y puesta en producción.

Resultados y Objetivos

Se han implementado diversos métodos (conocidos y también propios) de generación de CLs en el proyecto Java en GitHub.com [Gal15c]; algunos mejoran la eficiencia con respecto al método de J&M, aunque no obtienen CLs tan uniformemente distribuidos. Estas implementaciones son útiles cuando la uniformidad no es una condición que se requiera estrictamente. Podrían además utilizarse los distintos métodos alternativamente, como por ejemplo J&M para el inicio de la comunicación y el método del grafo de reemplazos en generaciones posteriores. De esta manera, se lograría una negociación entre tiempo y uniformidad de los resultados, aprovechando las ventajas de cada algoritmo.

Utilizando un concepto de producto de CLs de Koscielny, en [Gal13] se dio una implementación de un algoritmo muy eficiente ($O(n^2)$) para generar CLs de cualquier orden, aunque los resultados no son uniformes.

En el trabajo [Gal14] se da una implementación optimizada de J&M, que puede ejecutarse paso a paso para ver los resultados intermedios y finales mediante gráficos en 3D (utilizando la librería "OpenGL"). Esta propuesta se desarrolló con fines didácticos y educativos.

El método propuesto en [Gal15a] (no publicado), también detallado en el capítulo 5 de la tesis [Gal15d], converge en forma muy eficiente ($O(n^3)$ pero menor al algoritmo de J&M) a CLs prácticamente uniformes.

Por otra parte, el método de grafo de reemplazos publicado en [Gal15b], supera ampliamente en eficiencia (en un factor de 5 a 1) a la implementación dada del método de J&M, y la uniformidad de los CLs alcanzada es aceptable, lo cual lo hace útil en la práctica.

Los objetivos a largo plazo son investigar otras estructuras algebraicas como las propuestas por Koscielny (en [Kos95], [Kos02] y [Kos04]). Por ejemplo, los anillos, grupos, campos de Galois "adulterados" y otras estructuras son útiles para implementar soluciones eficientes en lenguajes de programación modernos como Java, C++ o Python. Esta línea se propone porque Koscielny da implementaciones en Maple 7, que es un lenguaje matemático muy específico, no disponible para sistemas que usen lenguajes de propósito general como los antes mencionados.

Además de generar CLs, se pretende estudiar la utilización de estas estructuras en aplicaciones prácticas (criptográficas o no), como por ejemplo en mecanismos de autenticación, algoritmos de cifrado/descifrado (alternativos al citado [Gib12]), protocolos de conocimiento cero, generadores de números pseudo-aleatorios (pseudo-random number generators o PRNGs), utilizar CLs como key-stream generators en stream ciphers (es decir, como generadores de claves), y

estudiar posibles ataques cuando se usan CLs en criptografía.

Formación de Recursos Humanos

Se prevé seguir trabajando sobre las líneas de investigación inconclusas y en nuevas líneas no consideradas hasta la actualidad. Se pretende publicar los resultados en congresos de seguridad informática o eventos especializados en criptografía a nivel nacional (como se hizo desde el inicio del proyecto). Si es posible, se desea también enviar trabajos a congresos o revistas especializadas a nivel internacional; el objetivo de esto último es interactuar con especialistas en criptografía para enriquecer el estudio. Si se consiguen resultados a nivel internacional, se considerará incluirlos en una posible tesis de doctorado en Ciencias de la Computación.

También sería muy conveniente conformar un equipo de trabajo que contribuya en las actividades de I+D+I en el tema propuesto. Se desea también buscar posibles formas de financiamiento del proyecto, lo cual permitiría brindar al equipo un feedback positivo en temas relacionados con la seguridad informática, y especialmente en criptografía.

Referencias

- [Gib12] Steve Gibson, *Off the grid* (Gibson Research online), <https://www.grc.com/offthegrid.htm> (2012).
- [JM96] Mark T. Jacobson and Peter Matthews, *Generating uniformly distributed random Latin squares*, J. Combin. Des. 4 (1996), no. 6, 405-437. MR MR1410617 (98b:05021)
- [MW91] Brendan D. McKay and Nicholas C. Wormald, *Uniform generation of random Latin rectangles*, J. Combin. Math. Combin. Comput. 9 (1991), 179-186. MR 1111853 (92b:05013).
- [MW05] Brendan D. McKay and Ian M. Wanless, *On the number of latin squares*, Ann. Combin. 9 (2005), 335-344.
- [SVT14] D. Selvi, G. Velammal, and Thevasahayam Arockiadoss, *Modified method of generating randomized Latin squares*, Journal of Computer Engineering (IOSR-JCE) 16 (2014), 76-80.
- [Kos95] Czeslaw Koscielny, *Spurious Galois fields*, Int. J. Appl. Math. Comput. Sci. 5 (1995), no. 1, 169-188.
- [Kos02] Czeslaw Koscielny, *Generating quasigroups for cryptographic applications*, Int. J. Appl. Math. Comput. Sci. 12 (2002), no. 4, 559-569.
- [Kos04] Czeslaw Koscielny, *Spurious multiplicative group of $GF(pm)$: a new tool for cryptography*, no. 12, 61-73.
- [Fon13] Fontana R.: *Random Latin squares and Sudoku designs generation*. ArXiv e-prints, 2013.
- [Car63] O'Carroll, F, *A Method of Generating Randomized Latin Squares*, Biometrics. December 1963, 652-653.
- [Gal13] Ignacio Gallego Sagastume, *Un método para la generación de cuadrados latinos de orden 256*, Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI). UTN Regional Córdoba, Argentina. (2013).

[Gal14] Ignacio Gallego Sagastume, *Generation of Latin squares step by step and graphically*, Congreso Nacional de Ciencias de la Computación (CACIC) 2014. Universidad de La Matanza, Buenos Aires, Argentina. (2014).

[Gal15a] Ignacio Gallego Sagastume, *Generation of Random Latin Squares Using a Random-Swapping Technique* (no publicado), 2015.

[Gal15b] Ignacio Gallego Sagastume, *Generación de cuadrados latinos de orden 256 utilizando un grafo de reemplazos*, Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNalISI). UTN Regional Buenos Aires, Argentina (2015).

[Gal15c] Ignacio Gallego Sagastume, *Proyecto "igs-lsgp" (latin square generation package) en github*, <https://github.com/bluemontag/igs-lsgp/wiki>, 2015.

[Gal15d] Ignacio Gallego Sagastume, tesis de Magister en Ing. De Software "*Análisis de algoritmos para generación de cuadrados Latinos aleatorios para criptografía*", agosto de 2015.
<http://sedici.unlp.edu.ar/handle/10915/475>
[77](#)

Ciberdefensa en Redes Industriales

Jorge Kamlofsky, Samira Abdel Masih, Hugo Colombo, Daniel Veiga, Pedro Hecht

CAETI - Universidad Abierta Interamericana

Av. Montes de Oca 725 – Buenos Aires – Argentina

(+54 11) 4301-5323; 4301-5240; 4301-5248

{Jorge.Kamlofsky, Samira.Abdel.Masih, Hugo.Colombo, Daniel.Veiga}@uai.edu.ar,
phecht@dc.uba.ar

Resumen

Los sistemas de control industrial han logrado la automatización de los procesos de producción en las industrias de manera robusta y eficiente. Gracias a ello, también fueron adoptados en infraestructuras críticas de las naciones: plantas potabilizadoras, distribución de energía, y demás. Nuevos requerimientos de mayor flexibilidad y eficiencia promueven su conexión con las redes corporativas, las que poseen gran cantidad de falencias y dejan expuestas a las redes industriales a sus vulnerabilidades y fallas.

En este proyecto se estudian las principales vulnerabilidades, ataques publicados, y se analizan y se desarrollan soluciones criptográficas basadas en el álgebra no conmutativa aptas para procesadores de menor porte, lo que permitirá otorgar seguridad criptográfica a los dispositivos de las redes industriales.

Palabras clave: Seguridad en Redes Industriales, Seguridad en Sistemas SCADA, Criptografía en PLCs, Ciberdefensa en SCADA, ciberdefensa en redes industriales.

Contexto

Los proyectos radicados en el CAETI¹ se clasifican en cinco líneas de investigación. Este proyecto se enmarca dentro la línea de investigación de “Seguridad Informática y Telecomunicaciones”. Se pretende obtener conocimiento teórico y desarrollar e implementar soluciones que permitan mejorar la situación de vulnerabilidad de estas redes.

Introducción

La disminución de los costos de chips y procesadores, así como también de las telecomunicaciones logró que los sistemas informáticos actuales se encuentren presentes en una cantidad enorme de lugares facilitando mejoras en productos y servicios. Su interconexión creciente a través de internet ha llegado incluso hasta el ámbito doméstico [1]. Surgen a diario, nuevas vulnerabilidades y ataques, gracias a la profesionalización del malware. Se requieren estrategias

¹Centro de Altos Estudios en Tecnología Informática, dependiente de la Facultad de Tecnología informática de la UAI.

claras y concisas de seguridad que minimicen riesgos con un esquema de defensa en profundidad [2].

El caso de los Sistemas de Control Industrial (ICS de sus siglas en inglés) es distinto. Los ICS son redes de tele-mando y tele-control de procesos compuestos por autómatas industriales (los “PLCs”) interconectados entre sí y cada uno de ellos a sensores (caudalímetros, sensores de nivel, de temperatura, etc.) y/o a actuadores (motores, válvulas, etc.). Se diseñaron para supervisar y actuar sobre procesos industriales.

Los ICS son muy robustos. Hoy están presentes en plantas de potabilización de agua, producción y distribución de energía, transporte, telecomunicaciones, es decir, están en infraestructuras críticas de naciones. Los SCADA (por sus siglas en inglés: Supervisory Control and Data Acquisition) se idearon para controlar sistemas industriales, conectando PC y redes de autómatas industriales; conformando la interfaz hombre máquina.

Con el tiempo surgió la necesidad de vincularlos a la red corporativa e incluso a internet. Y esta tendencia es creciente. Su interconexión dejó a los ICS expuestos a amenazas y riesgos, los que suponen serias consecuencias [3]. Hoy es posible, mediante dispositivos móviles, controlar un ICS, desde cualquier lugar del mundo con cobertura de red móvil [4], suponiendo un escenario ideal para explotar vulnerabilidades e inyectar malware.

La tecnología corporativa y la industrial dejaron a la seguridad entre ambas [5]. En la tecnología industrial, la seguridad carece de prioridad. Sí lo es el

proceso. El aislamiento de los procesos de producción industrial, le dio a los ICS una ilusoria sensación de seguridad por ocultamiento, durante muchos años [6, 7].

Hasta hace pocos años, era impensable que un ICS se pudiera infectar con virus informático. En el año 2010 el sistema SCADA de las plantas de enriquecimiento de uranio de Irán fue atacado por un virus llamado Stuxnet. Esto desconcertó a analistas estratégicos de todo el mundo. La comunidad internacional mostró preocupación por la seguridad [8 – 10], y se encuentra trabajando en soluciones [11 – 14].

En el ámbito de las tecnologías corporativas se tiene experiencia en Seguridad. Las recomendaciones de las normas ISO27000 ayudan a proteger la seguridad de los activos informáticos [15]. La criptografía es clave para asegurar sistemas informáticos. Es posible dar seguridad criptográfica a dispositivos con baja capacidad de cómputo gracias al desarrollo de algoritmia criptográfica de clave pública basada en estructuras algebraicas de anillos no conmutativos [16, 17] la cual a fecha actual es inmune a ataques cuánticos y esquemas simétricos compactos como el presentado en [18].

Este proyecto pretende desarrollar soluciones criptográficas basadas en Algebra no Conmutativa integrándolas con esquemas simétricos compactos para ser implementadas en las redes industriales.

Líneas de Investigación, Desarrollo e Innovación

El equipo de investigación trabaja en dos ramas: matemática-criptografía y redes y hardware.

La rama matemática-criptográfica trabaja estudiando las estructuras de anillos no conmutativos y no asociativos y su posibilidad de aplicarlos criptográficamente como sistema de intercambio seguro de claves. Las variantes generadas se programan y se las pone a prueba en ambientes de simulación controlados.

La rama de redes y hardware se encuentra estudiando los protocolos de comunicaciones intervinientes con la intención de lograr implementar los algoritmos generados en las redes. Se estudian las vulnerabilidades más frecuentes en estos sistemas.

Resultados y Objetivos

La rama matemática-criptográfica ha logrado implementar el protocolo presentado en [16] y ha llegado a una mejora en tiempos de ejecución usando cuaterniones [17].

La rama de Redes y Hardware analizó ataques a infraestructuras críticas publicados en [19] y ha propuesto un enfoque para disminuir los efectos de ciber ataques [20].

Se encuentra en proceso la instalación en laboratorio de una red industrial y un sistema SCADA. Con ello, se pretende replicar algunos ataques y probar las soluciones propuestas.

El objetivo final del proyecto es el desarrollo de soluciones de Seguridad que

puedan implementarse en las redes de los ICS. El problema en cuestión es crítico y se encuentra latente en toda la infraestructura industrial del mundo. Con el avance de este proyecto pueden obtenerse soluciones transferibles a la industria.

Formación de Recursos Humanos

El proyecto está dirigido por el Lic. Jorge Kamlofsky y la Dra. Samira Abdel Masih. Integran el proyecto el PhD. Hugo Colombo, el Lic. Daniel Veiga y el Dr. Juan Pedro Hecht.

El equipo se completa con los siguientes alumnos de la UAI: Juan Pedernera, Matías Sliafertas, Federico Arrieta, Oscar Hidalgo Izzi, Oscar Morales y Pablo Oviedo.

- Juan Pedernera, Matías Sliafertas y Federico Arrieta son alumnos que se encuentran promediando la carrera de Ingeniería en Sistemas. Juan Pedernera y Matías poseen experiencias en Seguridad de Redes. Federico Arrieta posee buen manejo en varios lenguajes de programación. Su participación en el proyecto le permitirá adquirir capacidades formales en investigación en redes industriales, para realizar su Trabajo Final de carrera.
- Oscar Hidalgo Izzi, Oscar Morales y Pablo Oviedo son alumnos de la Licenciatura en Matemática, próximos a concluir. Su participación en el proyecto le permitirá adquirir los conocimientos para el armado de sus Tesis de grado, las cuales se basan en los fundamentos

matemáticos de la algoritmia criptográfica que se desarrolla en el proyecto.

Referencias

- [1] Gustafson, S., and Sheth, A. *Web of Things*. Computing Now 7.3, 2014.
- [2] Jara H y Pacheco F. *Ethical Hacking 2.0*. Usershop, 2012.
- [3] Sánchez P. *Sistema de Gestión de la Ciberseguridad Industrial* [En línea]. Universidad de Oviedo, (2013). [Consulta: 11/02/15]. Disponible en: <<http://dspace.sheol.uniovi.es/dspace/bitstream/10651/17741/1/TFM%20-%20PABLO%20SANCHEZ.pdf>>.
- [4] Opto 22, *Press Release: Updates groov to Easily Connect Modbus/TCP Devices with Smartphones and Tablets* [En línea], (2015). Disponible en: <http://www.modbus.org/member_docs/OPTO22-Jan2015.pdf> [Consulta: 14/08/2015].
- [5] Carrasco Navarro, O. y Villalón Puerta, A. *Una visión global de la ciberseguridad de los sistemas de control*. Revista SIC: ciberseguridad, seguridad de la información y privacidad 106, (2013), pp. 52-55.
- [6] Courtois, N. *The dark side of security by obscurity, and Cloning MiFare Classic Rail and Building Passes Anywhere, Anytime*. IACR Cryptology ePrint. Archivo 2009: 137, 2009.
- [7] Menezes, A., Van Oorschot, P., and Vanstone, S. *Handbook of applied cryptography*. CRC press, 1996.
- [8] Veramendi, R. *Ataques a la Seguridad Informática y Telecomunicaciones en el Contexto Internacional*. Revista del Instituto de Estudios Internacionales IDEI-Bolivia, 45(2), (2012), pp. 4-11.
- [9] Vazquez, S. *Ciberseguridad en Paraguay*. III Conferencia Internacional y Taller de Ciberseguridad e Infraestructuras Críticas de Información, Buenos Aires, 2015.
- [10] Corvalan, F. *Seguridad de Infraestructuras Críticas: Visión desde la Ciberdefensa*. III Conferencia Internacional y Taller de Ciberseguridad e Infraestructuras Críticas de Información, Buenos Aires, 2015.
- [11] Blackmer, M. *Cibersecurity for Industrial Control Networks*. III Conferencia Internacional y Taller de Ciberseguridad e Infraestructuras Críticas de Información, Buenos Aires, 2015.
- [12] Simoes, P., Cruz, T., Proença, J. and Monteiro, E. *Honeypots especializados para Redes de Control Industrial*. VII CIBSI. Panamá, 2013.
- [13] Arias, D. *Seguridad en Redes Industriales*. Trabajo Final, Universidad de Buenos Aires, 2013.
- [14] Paredes, I. *La protección de infraestructuras críticas y ciberseguridad industrial*. Red seguridad: revista especializada en seguridad informática, protección de datos y comunicaciones 62, (2013), pp. 49.
- [15] ISOTools, *ISO 27001* [En línea], (2015). Disponible en: <<https://www.isotools.org/normas/riesgos-y-seguridad/iso-27001>>. [Fecha de consulta: 14 de Agosto de 2015].
- [16] Hecht J. *Un modelo compacto de criptografía asimétrica empleando anillos no conmutativos*. V CIBSI, Montevideo, 2009.
- [17] Kamlofsky J., Hecht J., Abdel Masih, S. Hidalgo Izzi, O. *A Diffie Hellman compact model over commutative rings using quaternions*. VIII CIBSI, Quito, 2015.
- [18] Castro Lechtaler, A., Cipriano, M., García, E., Liporace, J., Maiorano, A., y Malvacio, E.. *Model design for a reduced variant of a Trivium Type Stream Cipher*. Journal of Computer Science & Technology, 14.
- [19] Security Incidents Organization, *RISI: The Repository of Industrial Incidents* [En línea], (2015). Disponible en: <<http://www.risidata.com/Database>> [Consulta: 14/08/2015].
- [20] Kamlofsky J, Colombo H, Sliafertas M y Pedernera J, *Un Enfoque para Disminuir los Efectos de los Ciber-ataques a las Infraestructuras Críticas*. III CONAIISI, Buenos Aires, 2015.

MODELO DE ANÁLISIS RELATIVO A LA PROTECCIÓN DE DATOS PERSONALES PARA PROYECTOS DE CÓMPUTO EN LA NUBE

Juan Cruz González Allonca, Darío Piccirilli & Ma. Florencia Pollo-Cattaneo
Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS), Programa de Maestría en Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional.
Facultad Regional Buenos Aires
Medrano 951 (C1179AAQ) Ciudad Autónoma de Argentina. Buenos Aires Tel +54 11 4867-7511
{juanallonca, dariopiccirilli, flo.pollo}@gmail.com

Resumen

La información es el activo más importante de las organizaciones. Es por ello que asegurar la protección de los datos personales y la privacidad de la información durante su ciclo de vida es crucial a la hora de utilizar servicios de cómputo en la nube (Cloud Computing).

El esquema de cómputo en la nube ofrece beneficios como: flexibilidad, accesibilidad, autoservicio bajo demanda, escalabilidad, gestión de grandes volúmenes de datos. Sin embargo, estas ventajas, muchas veces no contemplan cuestiones críticas como la seguridad de la información y la privacidad de los datos almacenados.

Teniendo en cuenta estas falencias, es deseable definir un proceso de análisis que permita describir y evaluar la reglamentación vigente en el país referida a la protección de datos personales en proyectos de cómputo en la nube en el exterior del país. El proceso de análisis propuesto permitirá identificar y valorar el grado de cumplimiento con la normativa local, lo que facilitará la toma de decisiones informadas, basadas no solo en criterios técnicos o económicos, sino también regulatorios.

Palabras clave:

Cómputo en la nube, Privacidad, Cloud Computing, Protección de Datos Personales.

Contexto

La presente investigación se desarrolla dentro de una línea de trabajo incipiente del Grupo de Estudio de Metodologías para

Ingeniería en Software y Sistemas de Información (GEMIS), conformado por un conjunto de docentes y alumnos de la Facultad Buenos Aires de la Universidad Tecnológica Nacional (UTN-FRBA). El grupo GEMIS, se halla abocado a la búsqueda de la sistematización de cuerpos de conocimientos y promoción sobre el campo de la Ingeniería en Sistemas de Información y la Ingeniería en Software, sus aplicaciones y abordajes metodológicos en todo tipo de escenarios (convencionales y no convencionales).

En el marco de la UTN-FRBA, el equipo de investigadores ha trabajado desde las carreras de grado y posgrado de Ingeniería en Sistemas de Información, integrando entre sus miembros a docentes de grado y de posgrado, articulando los resultados de investigaciones con el desarrollo de Trabajos Finales de Carrera, Trabajos Finales de Especialidad y Tesis de Maestría.

Introducción

En los últimos años gran cantidad de empresas se ven atraídas por las ventajas técnicas y los bajos costos de mantenimiento que ofrece el esquema de cómputo en la nube (CN). Flexibilidad, accesibilidad, autoservicio bajo demanda, escalabilidad, gestión de grandes volúmenes de datos, son algunos de los beneficios que ofrece este esquema de CN. Sin embargo, estas ventajas, muchas veces no contemplan cuestiones críticas como la seguridad de la información y la privacidad de los datos almacenados [1].

Actualmente, la información es el activo más importante de las organizaciones. Es por ello que asegurar la privacidad de la información

durante su ciclo de vida es crucial a la hora de utilizar este tipo de servicios.

El desconocimiento o la no aplicación de la normativa vigente pueden transformarse tanto en pérdida de confianza o daño en la imagen de una empresa o perjuicio económico y hasta en responsabilidades jurídicas [2-4]. Las preocupaciones por estos inconvenientes por lo general son lo suficientemente importantes para algunas empresas y organizaciones, tanto que les llevan a evitar implementar sus sistemas en arquitecturas de CN.

Los principales obstáculos para la adopción de servicios CN [5] se concentran en tres cuestiones vinculadas a la localización de los datos: privacidad, confidencialidad y las relacionadas con la propiedad y los derechos de los datos en la nube. Por lo tanto, al momento de iniciar un proyecto de estas características, es determinante adecuarse a la normativa local y a su vez, analizar las cláusulas relativas a la seguridad de la información, especialmente a la protección de los datos personales. Si bien existe legislación aplicable que determina la extensión de responsabilidad, tanto del cliente, como del proveedor de servicios de CN, pocas organizaciones realizan un análisis previo en este sentido, donde se le permita al usuario conocer su nivel de riesgo y de cumplimiento normativo [6].

Por consiguiente, a partir del presente estudio se definirá un proceso de análisis que permita a las empresas u organismos locales, describir y evaluar la reglamentación vigente referida a la protección de datos personales en proyectos de cómputo en la nube con proveedores alojados en el exterior del país. Este proceso de análisis permitirá verificar el grado de cumplimiento con la normativa vigente, sumando el aspecto regulatorio a los análisis de viabilidad de un proyecto de CN.

A su vez, se promueve la implementación del estándar ISO/IEC 27018 [7], que puede contribuir a proporcionar confianza a los clientes de estos servicios respecto de la capacidad de cumplimiento normativo de los proveedores.

Líneas de Investigación, Desarrollo e Innovación

Para llevar a cabo el presente trabajo, se debe identificar el modelo de CN (nuevo paradigma en la forma de brindar servicios de cómputo por demanda [8]). Este modelo presenta un cambio importante en el paradigma computacional actual, la transformación de la infraestructura y las aplicaciones, desde un mundo claramente dominado y administrado por las organizaciones, a otro donde un tercero confiable y conocido le brinda servicios de infraestructura y uso de aplicaciones.

La Cloud Security Alliance (CSA) [9] describe cinco características esenciales en las que se evidencian similitudes y diferencias con las estrategias de computación tradicionales [10]:

1. Autoservicio por demanda.
2. Amplio acceso a la red
3. Reservas de recursos en común.
4. Rapidez y elasticidad.
5. Servicio supervisado.

Por otro lado, existen tres modelos distintos de prestación de los servicios en la nube [11]: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) y Software as a Service (SaaS)

Independientemente del modelo de servicio utilizado, existen cuatro formas de despliegue de los servicios de CN [12]: Nube Privada, Nube Pública, Nube Híbrida y Nube Comunitaria.

Estos modelos y formas de despliegue de los servicios de CN implican el tratamiento de datos de carácter personal [13], y su regulación que se encuentra determinada en el cuerpo jurídico local.

La protección de datos es un derecho a la intimidad personal que tienen las personas contra un tratamiento incorrecto, no autorizado o contrario a las normativas vigentes de sus datos personales por tratadores de datos. Al proteger los datos personales frente al riesgo de la recopilación y el mal uso, se ampara por ende, la privacidad de las personas.

La presente investigación, se centra en la Ley 25.326 [14], que establece las pautas para

la transferencia internacional de datos y la prestación de servicios informatizados de información en proyectos de CN. Es una norma de orden público que regula la actividad de las bases de datos que registran información de carácter personal y garantiza al titular de los datos la posibilidad de controlar el uso de sus datos.

La contratación de servicios de cómputo en la nube (que para la legislación argentina es una prestación de servicios informatizados) implica necesariamente un tratamiento de datos personales por terceros y las obligaciones de este tratamiento se encuentran determinadas su artículo 25 de la Ley 25.326 [14] y en el mismo artículo del Decreto 1558/01 [15], que reglamenta dicha ley.

Poner en marcha una estrategia de CN en el exterior del país implica necesariamente la transferencia internacional de datos de carácter personal. Esto genera que el control de la información deja de estar bajo el dominio del usuario y entra en la órbita de un tercero. La ley argentina pone particular atención en este tipo de tratamiento regulándolo específicamente, ya que los principios y derechos incluidos en la misma corren riesgos si no se establece un control que constituya límites de garantía y seguridad en la transferencia de los datos hacia otros países.

La cesión de datos personales dentro del país no sufre restricciones¹. Sin embargo, el panorama cambia al momento de transferir datos al exterior del país. En este último caso la ley contempla ciertos requisitos para que estas cuenten con garantías necesarias de respeto a la protección de la vida privada de los afectados y a sus derechos.

Para la ley 25.326, una transferencia internacional es un tipo de tratamiento de datos que consiste en la transmisión de datos, fuera de un Estado, realizado por el responsable del

¹ La transferencia internacional de datos personales está específicamente regulada por el art. 12 de la Ley N° 25.326, en el que se dispone la prohibición de transferir datos personales de cualquier tipo con países u organismos internacionales que no proporcionen niveles de protección adecuados. En la mencionada ley, no se imponen restricciones para la transferencia de datos dentro del país.

tratamiento a una persona física o jurídica, que los recibirá en un tercer país, para aplicarles un nuevo tratamiento, bien sea por cuenta propia o por cuenta del transmitente de los datos.

El principio general en materia de transferencia internacional se encuentra establecido en el artículo 12 de la Ley 25.326 [14] y dispone que la misma será lícita únicamente cuando el país importador de los datos tenga una legislación adecuada o equiparable a la del país exportador. Debido a la poca flexibilidad que otorga, se introdujo una excepción fundamental, a través del Decreto 1558/01, que establece que la prohibición de transferir datos personales hacia países u organismos que no proporcionen niveles de protección adecuados, no rige cuando el titular de los datos hubiera consentido expresamente o exista un contrato adecuado que garantice el nivel de protección de los datos personales.

El uso de servicios de computación en la nube, como se ha señalado, ofrece un gran número de ventajas. Pero a su vez presenta, por sus características específicas, una serie de riesgos que deben afrontarse con una adecuada gestión. En este aspecto las organizaciones deben estar atentas para revisar entre otros aspectos las obligaciones de cumplimiento regulatorio propias de la organización (como normas y procedimientos de seguridad corporativos) cómo a su vez la regulación local y de los países donde se procesarán los datos.

Resultados y Objetivos

Teniendo en cuenta las falencias presentes, se concluye que es deseable, definir un proceso de análisis que permita describir y evaluar la reglamentación vigente en el país referida a la protección de datos personales en proyectos de cómputo en la nube en el exterior del país. El proceso de análisis propuesto permitirá identificar y valorar el grado de cumplimiento con la normativa local, lo que facilitará la toma de decisiones informadas, basadas no solo en criterios técnicos o económicos, sino también regulatorios.

Se buscará la validación del proceso de análisis mediante el estudio de casos empíricos

de implementación de servicios de CN a través de proveedores en el exterior del país.

Para ello, se proponen los siguientes objetivos específicos:

- Promover el uso buenas prácticas de seguridad y privacidad en CN establecidas en el estándar ISO/IEC 27018 [7].
- Individualizar y analizar la legislación Argentina relativa a la transferencia internacional de datos y a la prestación de servicios informatizados de datos personales.
- Identificar los principales riesgos asociados a la los datos personales en el modelo de negocios de CN.
- Desarrollar una propuesta inicial de metodología de evaluación de riesgos en materia de protección de datos personales en servicios de CN.

Metodología de Trabajo

Las tareas a realizar para conseguir los objetivos antes mencionados serán las siguientes:

Para construir el conocimiento se seguirá un enfoque de investigación clásico [16-17] con énfasis en la producción de tecnologías [18], identificando métodos y materiales necesarios para desarrollar el proyecto.

Dentro de los métodos a utilizar se prevé: revisiones sistemáticas [19] y entrevistas con expertos [20]. Los expertos a entrevistar son: Dr. Pablo Palazzi [21], Dr. Horacio Azzolin [22], Iván Arce [23] y Dr. Daniel Altmark [24]. A continuación se detallan los materiales a utilizar:

- Ley N° 25.326 [14] y Decreto N° 1558/01 [15].
- Disposiciones y Dictámenes publicados por la Dirección Nacional de Protección de Datos Personales [25].
- Hemeroteca de la Universidad Tecnológica Nacional, Regional Buenos Aires [26].
- ISO/IEC 27018:2014 [6].
- Acceso a la biblioteca digital de IEEE [27].
- Acceso a la biblioteca digital de Springer [28].
- Acceso a la biblioteca digital de European Union Agency for Network and Information Security (ENISA) [29].

- Guía para la Evaluación de Impacto en la Protección de Datos Personales (Agencia Española de Protección de Datos) [30].

La presente investigación adopta una perspectiva descriptiva, empleando principalmente elementos cualitativos para identificar y describir las normas relativas a la protección de los datos personales que deben cumplirse en la implementación de proyectos de cómputo en la nube utilizando proveedores del exterior del país.

Para alcanzar los objetivos trazados se propone: (i) individualizar y analizar la legislación argentina relativa a la transferencia internacional de datos y a la prestación de servicios informatizados de datos personales, (ii) identificar casos de estudio, (iii) identificar los principales riesgos relativos a la seguridad de la información y a la privacidad de los datos en ambientes de CN, (iv) analizar los marcos regulatorios de otros países en cuanto a la transferencia internacional de datos personales, (v) realizar entrevistas semi estructuradas con expertos para validar la información recolectada en los puntos i, iii y iv, (vi) desarrollar una propuesta inicial de metodología de evaluación de riesgos en materia de protección de datos personales en servicios de CN, y (vii) validar los casos empíricos de implementación de servicios de CN en el exterior del país.

Formación de Recursos Humanos

El proyecto busca tanto la obtención de nuevos conocimientos como la motivación de los implicados para que asciendan dentro de la carrera de investigadores.

Los recursos humanos deben poseer una firme convicción que la protección de los datos personales y el resguardo de la privacidad y la intimidad son centrales para el desarrollo de las personas y las organizaciones. De esta manera, se logra un doble beneficio, el proyecto obtiene e incorpora el conocimiento regulatorio y técnico de los recursos humanos en el área de la especialidad, a la vez que plantea un esquema de formación de especialistas de punta en el proceso de gestión.

Finalmente, en el marco de este proyecto de investigación se encuentra radicada una Tesis de Magister en Ingeniería en Sistemas de Información y un Trabajo Final de Especialidad. Asimismo, se prevé incorporar alumnos avanzados en la carrera de Ingeniería en Sistemas de Información con posibilidades de articular sus Trabajos Finales de Carrera de Grado. De esta manera se espera generar un verdadero espacio integrado de investigación en los niveles de carreras de grado y posgrado.

Referencias

- [1]. European Union: European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Unleashing the Potential of Cloud Computing in Europe, 27 September 2012, COM(2012) 529 final, disponible en: <http://goo.gl/hBJKi7> disponible online en Marzo 2016.
- [2]. Blodget, H. (2011), Amazon's Cloud Crash Disaster Permanently Destroyed Many Customers' Data <http://goo.gl/CsDRKz> disponible online en Marzo 2016.
- [3]. Wittow, M. H. (2011) Cloud Computing: Recent Cases and Anticipating New Types of Claims <http://goo.gl/PXEGKr> disponible online en Marzo 2016.
- [4]. Sherman, M. (2014) Supreme Court Justices concerned over impact on cloud computing in Aereo case <http://goo.gl/TSVh4J>. Disponible online en Marzo 2016
- [5]. Etro, F. (2010) "The Economic Consequences of the Diffusion of Cloud Computing" en Dutta, Soumitra; Mia, Irene. The Global Information Technology Report 2009 – 2010 ICT for Sustainability. Londres, Foro Económico Mundial - INSEAD.
- [6]. González Allonca J. C., Ruiz Martínez E. (2015) Cloud Computing: la regulación de la transferencia internacional de datos personales y la prestación de servicios por parte de terceros, Ed. Infojus DACF150527. <http://goo.gl/efNsq6> disponible online en Marzo 2016.
- [7]. ISO/IEC 27018:2014. (2014) Information technology – Security techniques - Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors, International Organization for Standardization ISO.
- [8]. Ravindran, A. (2013) Eemerging cloud computing paradigm and its impact on enterprises. Business Review: Advanced Applications, Cambridge Scholars Publishing <http://goo.gl/JAPTjR> disponible online en Marzo 2016.
- [9]. Cloud Security Alliance, Security Guidance for Critical Areas of Focus in Cloud Computing. Versión 3.0 (2011) <https://goo.gl/cvCDgm> disponible online en Marzo 2016
- [10]. Ludwig S., 'Cloud 101: What the heck do IaaS, PaaS and SaaS companies do?', <http://goo.gl/kn91F5> disponible online en Marzo 2016.
- [11]. Voorsluys, W.; Broberg, J.; Buyya, R. (2011) Introduction to Cloud Computing. En R. Buyya, J. Broberg, A.Goscinski. Cloud Computing: Principles and Paradigms . EEUU: Wiley pp. 1–44. ISBN 978-0-470-88799-8.
- [12]. McFedries, P. (2008) The Cloud Is The Computer. IEEE Spectrum Magazine <http://goo.gl/GgQhHY> disponible online en Marzo 2016.
- [13]. Becerra, M., Navarro, M. (2012) Retos actuales para la protección de datos personales en las organizaciones. Universidad Nacional de San Juan, FFCE y N , Proyecto Código N°21/E/871. "Convergencia de Tecnologías informáticas y Metodologías para la implementación de sistemas de Información". (178-192).
- [14]. Ley N° 25.326 (2000) Ley de Protección de Datos Personales. Boletín Oficial de la República Argentina.
- [15]. Decreto 1558 (2001) Reglamentación de la Ley N° 25.326. Boletín Oficial de la República Argentina.
- [16]. Riveros, H. y Rosas, L. (1985) El Método Científico Aplicado a las Ciencias Experimentales. México: Editorial Trillas. ISBN 96-8243-893-4.
- [17]. Creswell, J. (2002) Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research. Prentice Hall. ISBN 10: 01-3613-550-1.
- [18]. Sabato J, Mackenzie M. (1982). La Producción de Tecnología: Autónoma o Transnacional. Instituto Latinoamericano de Estudios Transnacionales - Technology & Engineering. ISBN 9789684293489.
- [19]. Argimón J. (2004). Métodos de Investigación Clínica y Epidemiológica. Elsevier España, S.A. ISBN 9788481747096.
- [20]. Martínez, M. (2006) La Investigación Cualitativa (Síntesis conceptual). Revista de investigación en psicología, ISSN-e 1560-909X, Vol. 9, N°. 1, 2006 , pp. 123-146
- [21]. Dr. Pablo Palazzi, Curriculum Vitae disponible en: <http://goo.gl/VNe0nM>
- [22]. Dr. Horacio Azzolin, Curriculum Vitae disponible en: <https://goo.gl/gvhMpw>
- [23]. Iván Arce, Curriculum Vitae disponible en: <https://ar.linkedin.com/in/ivanarce>
- [24]. Dr. Daniel Altmark, Curriculum Vitae disponible en: <http://goo.gl/YO8s8Y>
- [25]. Acceso a normativa publicada por la Dirección Nacional de Protección de Datos Personales <http://www.jus.gob.ar/datos-personales> disponible online en Marzo 2016.
- [26]. Hemeroteca de la Universidad Tecnológica Nacional, Regional Buenos Aires, Escuela de Posgrado. <http://goo.gl/taoMv2>
- [27]. Biblioteca digital de IEEE <http://goo.gl/uICeGj> disponible online en Marzo 2016.
- [28]. Biblioteca digital de Springer <http://goo.gl/svKQj4> disponible online en Marzo 2016.
- [29]. Biblioteca digital de European Union Agency for Network and Information Security (ENISA) <https://goo.gl/Z0UZBL> disponible online en Marzo 2016.
- [30]. Agencia Española de Protección de Datos (2014) Guía para la Evaluación de Impacto en la Protección de Datos Personales, Agencia Española de Protección de Datos. <https://goo.gl/FZTj8v> disponible online en Marzo 2016.

Análisis digital forense utilizando herramientas de software libre

Lic. Francisco Javier Díaz - Lic. Paula Venosa - Lic. Nicolás Macia - Lic. Einar Lanfranco
Lic. Alejandro Sabolansky - Damian Rubio

[javierd | pvenosa | nmacia | einar | asabolansky | drubio] at linti.unlp.edu.ar

LINTI (Laboratorio de Investigación en Nuevas Tecnologías Informáticas)
Facultad de Informática - UNLP
Calle 50 y 120 – 2do piso – La Plata, Buenos Aires, Argentina

1. Resumen

En la actualidad el mundo es digital. La mayor parte de la población utiliza medios digitales. Hoy en día se conecta a Internet gran cantidad de hardware y software sin considerar, muchas veces, los problemas de seguridad asociados con su uso.

Cuando se aprovecha una vulnerabilidad de un sistema informático y se concreta la amenaza mediante un ataque, nos encontramos con un incidente de seguridad que hay que estudiar.

En cualquier caso se necesita saber qué fue lo que sucedió teniendo como principal objetivo determinar las causas del problema para poder solucionarlo e identificar a los responsables del mismo. Ante esto, el análisis forense digital toma una gran relevancia.

Entre los alcances esperados de esta línea de I/D/I se espera ganar experiencia en lo que se refiere al campo de investigación del estudio digital forense. Para ello, se pretende identificar y probar herramientas de software libre que puedan ser utilizadas cuando se realiza este tipo de actividad. Además, en base al conocimiento adquirido, se espera generar documentos de buenas prácticas, procedimientos y materiales adecuados que puedan ser utilizados en cursos y capacitaciones sobre temáticas relacionadas con el estudio forense digital.

Palabras clave: seguridad de la información, forensia digital, software libre

2. Contexto

En el Laboratorio de Investigación en Nuevas Tecnologías Informáticas (LINTI) [1] de la Facultad de

informática de la Universidad Nacional de La Plata [2], un grupo de docentes/investigadores se dedica a estudiar temas relacionados con la seguridad y privacidad de la información, aplicando los conocimientos en los distintos proyectos en el que participan.

En el marco del proyecto de incentivos “Internet del Futuro: Ciudades Digitales Inclusivas, Innovadoras y Sustentables, IoT, Ciberseguridad, Espacios de Aprendizaje del Futuro”, este grupo investiga vulnerabilidades de seguridad actuales que afectan a sistemas, redes y servicios. En particular, la línea que se presenta se enfoca en el estudio de la forensia digital en los incidentes de seguridad, utilizando software libre, considerando:

- Relevamiento de buenas prácticas en el campo de la investigación forense.
- Definición de documento de procedimientos a aplicar.
- Relevamiento de las herramientas existentes para cada uno de los pasos del procedimiento.
- Relevamiento de los dispositivos de hardware que sean elementales.
- Elección de las herramientas que se consideren adecuadas.
- Generación de guía de análisis forense para capacitar.

Este grupo de investigadores forma parte del Centro de excelencia de la UNLP en el tema “Ciberseguridad” [3], seleccionado por la UIT (Unión internacional de Telecomunicaciones) para el año 2015.

3. Introducción

En la actualidad el mundo es digital, la mayor parte de la población utiliza medios digitales. Si bien esta adopción fue en paulatino y constante crecimiento desde la aparición de las primeras computadoras personales allá por el año 1981 [6], en los últimos años se aceleró de manera notable la penetración de la tecnología en el día a día de las personas. Esto se debió a distintos factores como:

- La disponibilidad de conexión a Internet, sobre todo con la masificación del acceso a través de banda ancha.
- La aparición de dispositivos móviles, en particular los denominados smartphones o teléfonos inteligentes [7].
- La llegada del fenómeno denominado Internet de las Cosas (IoT) [8].

La aparición de IoT ha generado un crecimiento exponencial de la cantidad de dispositivos conectados por persona. Por ejemplo, la persona que antes tenía una PC, hoy suele tener en el mejor de los casos la PC, un router inalámbrico, el celular, la tablet y el TV. Esta evolución de dispositivos interconectados va de la mano de la funcionalidad que brindan generando que se desarrolle y conecte muchísimo hardware y software sin considerar, muchas veces, los problemas de seguridad que su uso trae asociado. Esto se debe a que se desarrollan dispositivos pensando en la funcionalidad y usabilidad de los productos y no en la seguridad de la información que los mismos manipulan [9].

Cuando a través de un ataque informático se concreta una amenaza aprovechando una vulnerabilidad existente, se está frente a un incidente de seguridad. Los incidentes de seguridad deben ser analizados, ya sea tanto para solucionar el problema como para poder determinar el origen del mismo. Esta investigación se la conoce como análisis forense. El análisis forense digital, según Miguel López Delgado [10], en un sentido formal, es definido como un conjunto de principios y técnicas que comprende el proceso de adquisición, conservación, documentación, análisis y presentación de evidencias digitales y que llegado el caso puedan ser aceptadas legalmente en un proceso judicial.

El análisis forense se puede realizar simplemente a nivel institucional o llegar al ámbito de la justicia civil o penal. A nivel mundial la legislación se está adecuando a los nuevos tiempos que vive el mundo [11] [12], donde la mayoría de los delitos que existían en el mundo no digital se trasladan al mundo virtual. Es importante destacar que en muchos países el fraude electrónico se encuentra en el podio de los delitos más efectivos [13], considerándose muchas veces tan rentable como el narcotráfico.

Aún en instituciones donde las pérdidas no son económicas, ante un ataque exitoso aparecen problemas de diversos tipos, como ser en el caso de la UNLP donde el defacement de la página institucional implica una pérdida de prestigio [14] o la pérdida de disponibilidad de los recursos de la organización, por ejemplo.

Cualquiera sea el perjuicio producido se necesita saber qué fue lo que pasó, para paliar la situación, para solucionarla y para perseguir a los responsables. Para lograrlo es necesario el análisis forense y de allí su importancia.

A esta disciplina, se la puede definir como un área perteneciente al ámbito de la seguridad informática surgida a raíz del incremento de los diferentes incidentes de seguridad. Éste es un tipo de análisis que se realiza con posterioridad a la ocurrencia de los incidentes, intentando reconstruir cómo se ha penetrado o vulnerado un sistema y en lo posible identificando a los responsables.

A modo de resumen, podemos decir que se intenta responder al menos a las siguientes preguntas:

- ¿Quién ha realizado el ataque?
- ¿Cuándo y cómo se hizo?
- ¿Qué vulnerabilidades aprovechó?
- ¿Qué hizo el atacante una vez que tuvo éxito y accedió al sistema?

Para llevar adelante estos estudios existen una serie de herramientas de hardware y de software que posibilitan tanto llevarlos a cabo técnicamente como que sus resultados tengan validez, ya que hay que tener en cuenta una serie de cuestiones que tienen que ver con la preservación de la evidencia. Un ejemplo de este tipo de herramientas es un dispositivo, denominado duplicador [15], el cuál permite realizar una copia exacta de la información de un

disco rígido a nivel físico sin alterar la información original existente en el dispositivo a analizar.

4. Líneas de Investigación, Desarrollo e Innovación

Sobre los ejes de investigación, inicialmente planteados: forensia digital utilizando software libre hasta el momento podemos mencionar que:

- Realizamos una recopilación de bibliografía para conocer el estado del arte actual.
- Participamos de una serie de charlas en distintas conferencias de seguridad que trataban la temática.
- Asistimos al cyberdrill para CSIRTS realizado por la ITU .
- Participamos de seminarios web dictados por la OEA.
- En el marco del trabajo diario de CERTUNLP hemos realizado varias forensias digitales ante eventos.

5. Resultados y Objetivos

5.1. Objetivo General

Se espera ganar experiencia en lo que se refiere al campo de investigación del estudio digital forense, lo que permitirá hacer una selección de un conjunto de herramientas de software libre que puedan ser recomendadas para realizar este tipo de actividad.

5.2. Objetivos Específicos

- Indagar sobre las prácticas recomendadas en forensia digital.
- Identificar los posibles escenarios de un estudio forense y las mejores técnicas a seguir en cada uno de ellos.
- Relevar las múltiples herramientas de software existentes, poniendo especial énfasis en aquellas que sean de software libre.
- Relevar las múltiples herramientas de hardware existentes.

- Generar una lista de todas las herramientas que se consideren recomendables para este tipo de investigaciones.
- Generación de material que pueda ser utilizado en cursos de capacitación o cátedras de la Facultad relacionadas con la temática.
- Formar RRHH que retroalimenten al grupo de investigadores convirtiéndolo en un referente en el tema.

6. Formación de Recursos Humanos

La línea de investigación en forensia digital está siendo abordada por integrantes del LINTI que forman parte del grupo de seguridad: Nicolás Macia, Paula Venosa, Einar Lanfranco y Alejandro Sabolansky, quienes desarrollan su actividad también desde el año 2008 en CERTUNLP, el centro de respuesta a incidentes académico de la Universidad Nacional de La Plata.

En CERTUNLP trabajan tres becarios, entre ellos Damián Rubio, que se encuentra actualmente participando en las tareas relacionadas a forensia digital, formándose y capacitándose en esta temática.

El plan de trabajo del docente Lic. Einar Lanfranco se encuentra en sintonía con la presente línea de investigación, dirigido por el Lic. Javier Díaz y la Lic. Paula Venosa.

El grupo de seguridad del LINTI de la Facultad de Informática de la UNLP trabaja desde el año 2000 con distintas experiencias relacionadas con la Seguridad de la Información: auditorias de seguridad, implementación de infraestructuras de seguridad, consultoría, desarrollo e implementación de Software Libre y concientización.

Cabe destacar que este grupo de investigadores representa a la UNLP en el Centro de excelencia en el tema “Ciberseguridad” de la UIT, durante el transcurso del año 2015 [5].

Dentro de este marco, durante año 2016 se dictará el curso “Tratamiento y resultados desde la Evidencia digital usando software libre”, coordinado por el Lic. Einar Lanfranco.

Además el grupo comenzó a participar en el año 2014 de la comisión de estudio ITU-T SG17:Security

de la UIT [4], donde se abordan temas actuales de seguridad de la información.

La Facultad de Informática aprobó a fines de 2015 el proyecto de transferencia “La Forensia Digital en el mundo del software libre” coordinado por los docentes: Lic. Einar Lanfranco, Lic. Nicolas Macia y Lic. Paula Venosa, el cuál debe llevarse a cabo durante el año 2016.

Referencias

- [1] Laboratorio de Investigación de Nuevas Tecnologías Informáticas - LINTI.Facultad de Informática: <https://www.linti.unlp.edu.ar>
- [2] Facultad de Informática: <https://info.unlp.edu.ar>
- [3] Centro de excelencia en Ciberseguridad. <http://www.itu.int/en/ITU-D/Regional-Presence/Americas/Documents/EVENTS/2015/0225-BR-COE/Agenda-EN.pdf>
- [4] Grupo de Estudio 17 en la ITU <http://www.itu.int/en/ITU-T/studygroups/2013-2016/17/Pages/default.aspx>.
- [5] ITU. <http://www.itu.int>
- [6] http://www-03.ibm.com/ibm/history/exhibits/pc25/pc25_birth.html
- [7] <https://www.itu.int/net/itunews/issues/2011/03/12-es.aspx>
- [8] <http://www.cisco.com/web/solutions/trends/iot/overview.html>
- [9] <http://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/>
- [10] Análisis Forense Digital. Miguel López Delgado. http://www.oas.org/juridico/spanish/cyb_analisis_foren.pdf
- [11] http://www.coe.int/t/dg1/legalcooperation/economiccrime/cybercrime/Default_en.asp
- [12] <http://www.itu.int/en/ITU-D/Cybersecurity/Pages/Legal-Measures.aspx>
- [13] <http://www.lexisnexis.com/risk/downloads/assets/true-cost-fraud-2014.pdf>
- [14] <http://www.zone-h.org/archive/filter=1/fulltext=1/domain=unlp.edu.ar>
- [15] <https://www2.guidancesoftware.com/products/Pages/tableau/products/forensic-duplicators/td3.aspx>

Una evaluación de los simuladores de autómatas existentes a los fines de su implementación práctica en la materia Sintaxis y Semántica de los lenguajes en la carrera de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Córdoba

OLARIAGA, Sandra Mónica (1), PAEZ Nancy del Valle (2)

Dpto. Ingeniería en Sistemas de Información

Facultad Regional Córdoba

Universidad Tecnológica Nacional

Maestro Marcelo López s/n. Ciudad Universitaria.

Teléfono: 0351 – 5986000

solariaga@sistemas.frc.utn.edu.ar (1) / npaez@sistemas.frc.utn.edu.ar (2)

Resumen

Un autómata finito es un modelo matemático abstracto compuesto por una cantidad finita de estados que tiene como objetivo recibir entradas y generar salidas este es el caso del autómata transformador, si el autómata sólo se limita a reconocer cadenas el mismo estará conformado por un alfabeto de entrada, un conjunto de estados finito, una función de transición, un estado inicial y un conjunto de estados finales. Su funcionamiento se basa en una función de transición, que recibe a partir de un estado inicial una cadena de símbolos del alfabeto de entrada, va leyendo dicha cadena y el autómata se desplaza de un estado a otro, para finalmente detenerse en un estado de aceptación o de rechazo.

Dentro de los autómatas finitos, sean transformadores o reconocedores, se destacan el autómata finito determinista y el no determinista, la diferencia entre ambos radica en la función de transición, en la cual para el autómata determinista a cada estado le corresponde una única transición, es decir que para un estado determinado y ante un símbolo leído

existe siempre una sola transición posible, mientras que en el autómata finito no determinista existen múltiples transiciones posibles desde un estado y un símbolo de entrada.

Palabras clave: Simuladores, Autómatas finitos, Sintaxis y Semántica de los Lenguajes

Contexto

El Código del proyecto es UTN 3587 dentro del área de Ciencias de la Computación e Informática, radicado dentro de la Secretaria de Ciencia y Tecnología de la Universidad Tecnológica Nacional Facultad Regional Córdoba. Fecha de inicio: 01/05/2015 Fecha de finalización: 31/04/2017

Introducción

En el presente trabajo se realizará una evaluación de los simuladores de autómatas finitos que constituyen una herramienta muy útil a la hora de facilitar

la comprensión del funcionamiento de los autómatas finitos.

Para ubicarnos en el tipo de software que vamos a evaluar se trata de un software de simulación de entidades abstractas (autómatas) pero también de ejercitación ya que mediante el manejo del software contribuye a que los alumnos comprendan el funcionamiento de los autómatas a través de la retroalimentación y verificación del aprendizaje.

Varias universidades a nivel internacional han desarrollado simuladores de autómatas finitos a fin de poder ejecutar mediante los mismos la ejercitación de los alumnos, a saber:

* JFLAP Universidad de Duke. EEUU

*THOTH Universidad de Burgos. España

*Automaton Simulator Dr. Carl Burch Software Engineer, Google, Kirkland, Wash. (was Associate Professor of Computer Science, Hendrix College until Sep 2014) PhD, Computer Science, Carnegie Mellon U, 2000

* Minerva de la Universidad Nacional de Buenos Aires

* Visual Automata Simulator Computer Science University of San Francisco. EEUU

Esta evaluación de los simuladores de autómatas existentes pretende determinar la jerarquía de los mismos en cuanto a su potencial implementación en las prácticas de la materia Sintaxis y Semántica de los Lenguajes de la carrera de Ingeniería en Sistemas de Información.

Líneas de Investigación, Desarrollo e Innovación

El proyecto se inscribe dentro de los lineamientos de investigación en innovaciones curriculares en Educación Superior Universitaria dentro del área de Sistemas e interfaces adaptables en la intervención humano-computadora.

Se trata de una investigación descriptiva porque busca profundizar la comprensión del problema y está basada en la observación. La preocupación estará centrada en indagar los hechos o fenómenos educativos en su “realidad natural” . Se aplicarán las técnicas de la investigación acción. Para ello se seguirán los siguientes pasos:

El primer paso consiste en la identificación de la problemática acerca de la necesidad de evaluar los simuladores de autómatas finitos existentes a fin de poder determinar jerarquía de los mismos en cuanto a su potencial implementación en las prácticas de la materia Sintaxis y Semántica de Lenguajes de la carrera de Ingeniería en Sistemas de Información

El segundo paso es la investigación bibliográfica y por internet acerca de los simuladores de autómatas finitos existentes.

El tercer paso es la determinación de los parámetros a partir de los que se realizará el análisis de los simuladores encontrados.

El cuarto paso es el análisis del simulador en función de los parámetros establecidos.

El quinto paso es la documentación de los resultados obtenidos en base a los registros realizados.

Resultados y Objetivos

Objetivo general

Llevar a cabo el análisis de los simuladores de autómatas existentes a fin de determinar la jerarquía de los mismos considerando su potencial implementación en las prácticas de la materia Sintaxis y Semántica de Lenguajes de la carrera de Ingeniería en Sistemas de Información.

Objetivos específicos

- Llevar a cabo un relevamiento acerca de los simuladores de autómatas existentes.
- Determinar los parámetros de análisis tanto técnicos como pedagógicos sobre los cuales se procederá a la evaluación de los mismos.
- Efectuar la medición de cada uno de los simuladores encontrados de acuerdo a los parámetros establecidos
- Confeccionar un informe acerca de los aspectos evaluados
- Elaborar un informe acerca de la importancia de la implementación de los simuladores dentro de la práctica de la materia Sintaxis y Semántica de Lenguajes de la carrera de Ingeniería en Sistemas de Información

El estado actual de avance es el siguiente: nuestro trabajo se viene desarrollando a través de varias tareas, las cuales algunas de ellas se han realizado simultáneamente, dando el punto de partida a otras; a saber:

- Recabamos información acerca de los aspectos que se consideran para llevar a cabo la evaluación de software, tanto los técnicos como los pedagógicos.
- Discusión entre los integrantes del grupo acerca de cuáles serían los aspectos tanto técnicos como pedagógicos relevantes a ser incluidos en las tablas de evaluación de simuladores de autómatas.

Estamos abocados a la elaboración de una matriz de evaluación de acuerdo a la información recabada. Con la convicción de que no existe un patrón fijo para evaluar los aspectos técnicos y pedagógicos de un software, sino guías que permiten describir diferentes aspectos expresando usos deseables, pues tanto la calidad del software como la de cualquier otro material didáctico está relacionada y determinada en función de un docente concreto y de un grupo puntual de alumnos. Los criterios de evaluación propuestos ofrecen orientaciones, una descripción del uso posible y pedagógicamente viable; más que una evaluación formal.

Formación de Recursos Humanos

La Directora y la Co-directora de este PID son Especialistas en Docencia Universitaria, siendo la Directora Magister en Docencia Universitaria.

Con este proyecto se continuará con el proceso de formación en investigación de todos los integrantes, ya que los mismos han participado en proyectos anteriores, con temáticas referidas al ámbito de la educación y de competencias.

Este estudio en particular contribuirá en gran medida a ampliar el conocimiento acerca de la posible aplicación de simuladores de autómatas finitos en la práctica de la materia Sintaxis y Semántica de los lenguajes a fin de que pueda llegar a ser el punto de partida de actividades que favorezcan el mejoramiento de los procesos de enseñanza y de aprendizaje en el aula universitaria.

Se han incorporado becarios alumnos y graduados algunos de los cuales ya han participado en proyectos anteriores, de la carrera de Ingeniería en Sistemas de Información.

Referencias

* Alfonseca E., Alfonseca M. y Moriyón R. (2007): Teoría de Autómatas y Lenguajes Formales McGraw Hill.

* Garcia P., Perez T., Ruiz J., Segarra E. Sempere J. Vázquez de Parga, M.(2001) Teoría de Autómatas y Lenguajes Formales. Alfaomega.

* Gonzalez Castañon M. A. (2000) Evaluación de Software Educativo: orientaciones para su uso Universidad EAFIT Proyecto Conexiones.

* Hopcroft J., Ullman J. Motwani R. (2008) Teoría de autómatas, lenguajes y computación. Pearson Addison Wesley.

* Martin J. (2004) Lenguajes Formales y Teoría de la computación McGraw Hill.

* Sicardi, I. (2004). Revista de Informática Educativa y Medios Audiovisuales *Vol. 1(3)*, págs. 1-20. 2004 ISSN 1667-8338 © LIE-FI-UBA.

Herramientas Colaborativas Multiplataforma en la Enseñanza de la Ingeniería de Software

Carlos Neil, Marcelo De Vincenzi, Nicolas Battaglia, Roxana Martínez

CAETI – Centro de Altos Estudios en Tecnología Informática
Facultad de Tecnología Informática - Universidad Abierta Interamericana
{carlos.neil, medevincenzi, nicolas.battaglia, roxana.martinez}@uai.edu.ar

Resumen

En el uso de herramientas CASE (*Computer Aided Software Engineering*) como soporte de modelado para la enseñanza de la ingeniería de software, suelen utilizarse productos comerciales de uso profesional o herramientas de modelado de acceso libre no tan comunes en el mercado. Surge entonces la necesidad de contar con herramientas CASE que permitan asistir al alumno a adquirir y afianzar los contenidos académicos desarrollados dentro de un espacio que facilite tanto a sus pares como al docente colaborar entre sí. Se propone, en este trabajo de investigación y desarrollo, la integración de plataformas de trabajo colaborativo y plataformas de modelado UML, dentro de un entorno académico mixto, en donde la herramienta no reemplace al modelo presencial sino que brinde un espacio de colaboración que permita optimizar tanto las relaciones entre docentes y alumnos, docentes y docentes, alumnos y alumnos. La herramienta está pensada, además, para ser usada en la evaluación y seguimiento de los trabajos desarrollados por cada alumno a lo largo de su carrera. También permitirá la identificación de los

equipos de trabajo en los que participó cada alumno en su trayecto académico. El objetivo general planteado es el diseño e implementación de una herramienta CASE colaborativa multiplataforma, denominada *UAICase*.

Palabras clave: Trabajo Colaborativo, Ingeniería de Software, Herramientas de Desarrollo, Multiplataforma.

Contexto

Este proyecto enmarca y potencia el trabajo de integración curricular desarrollado en la facultad de Tecnología Informática de la UAI. A partir del uso de la herramienta CASE colaborativa (*UAICase*) los alumnos podrán adquirir y utilizar los conceptos de modelado e ingeniería de software en un proyecto áulico colaborativo que evolucionará, iterativa e incrementalmente, desde 2° año¹, en las asignaturas que introducen los conceptos de modelado, hasta 3° año² y 5° año³ donde desarrollan trabajos de

¹ Metodología de Desarrollo de Sistemas I y II.

² Base de Datos, Trabajo de Diploma y Trabajo de Campo I.

³ Seminario de Aplicación Profesional y Trabajo Final de Ingeniería.

integración de contenidos que incluyen las asignaturas de cierre del ciclo intermedio (analistas de sistemas) y del ciclo final (Ingeniero en Sistemas Informáticos).

Introducción

La presencia generalizada de computadoras y dispositivos que permiten tener información, en cualquier momento y en cualquier lugar demuestra de forma sustentable que la informática se ha vuelto ubicua (10). Estos avances en la tecnología crean nuevos contextos de comunicación e interacción entre usuarios, generando nuevos conjuntos de herramientas que reemplazan a otros más antiguas y aburridas (1). La capacidad de ubicuidad que genera este avance tecnológico, permite disponer de contenidos a lo largo del día, en cualquier momento y en cualquier lugar; obligando a disponer de nuevas infraestructuras tecnológicas (11).

Los avances en Tecnología de Información y Comunicación (TIC), generan grandes aportes a la educación, facilitando el acceso a la información en determinados contextos y con determinadas características (1) mediante herramientas que permiten acceder a la misma, de forma sincrónica o asincrónica (6), mediante el uso de *Chats*, Foros, Correos y *Blogs* que promueven la interacción entre personas (13), en este ámbito se modifican los aspectos formales de la educación presencial y se agregan otros nuevos como la educación en línea, ampliando el horizonte más allá de los límites físicos que impone la educación presencial (ambientes mixtos), en donde cada uno de los integrantes, aporta conocimientos, estilos y modos de aprender (19). Todos estos nuevos contextos de la educación aparecen

definidos como “la sociedad de la educación” y la “sociedad del conocimiento”, caracterizadas por la movilidad, interactividad y ubicuidad; ofreciendo, además, posibilidades constantes de aprendizaje dentro y fuera del aula (10). Dentro de las instituciones (y en particular en las carreras de ingeniería), los docentes apoyan las nuevas posibilidades de comunicación como apoyo para la organización y el dictado de sus materias, generando un gran impacto en la construcción colaborativa del conocimiento (15).

Es fundamental, en los ámbitos de formación, el papel que desarrolla el docente y qué propuesta haga sobre el uso de los medios de comunicación y la evolución tecnológica dentro del aula. Existe una rivalidad entre los conocimientos adquiridos fuera del aula con medios más modernos y llamativos, con los adquiridos dentro en donde, probablemente, sean menos atractivos y más aburridos (1). La educación en línea, aun siendo parte esencial de la educación a distancia, también suele ser utilizada como complemento de la educación presencial, la cual se califica como híbrida, bimodal o mixta.

Son las TIC un medio eficaz para avanzar hacia la redefinición de nuevas estrategias y características necesarias que permitan integrar al modelo presencial las tecnologías de educación online colaborativa, siendo las entidades educativas quienes tienen el desafío de implementar nuevas herramientas que puedan coexistir con los modelos clásicos presenciales (11)

Existen dos enfoques de educación: centrado en el docente o centrado en el alumno, y se identificó que ninguno de estos enfoques es completamente fiel a las líneas actuales de investigación sobre aprendizaje (4), tanto en ambientes presenciales como en línea, debido a que

no se tiene en cuenta la actividad mental del alumno en conjunto con el tutor ni principalmente entre alumnos.

En el aprendizaje colaborativo la participación y los intercambios entre pares en las aulas, tanto físicas como virtuales, se producen preferentemente a través de los diálogos; la discusión y el intercambio se constituyen en escenarios de aprendizaje en donde se requiere la comprensión conjunta de lo que se aborda, esto implica que cada uno de los que participan tenga en cuenta lo que el otro comprende, para construir una visión compartida entre todos (19).

Las plataformas de formación online están comenzando a disponer de herramientas de interacción en tiempo real entre el docente y el alumno, pero no están pensadas como complemento para el tipo de cursada presencial, por este motivo se plantea una solución tecnológica que utiliza las ventajas de cada uno de los puntos anteriores para que el impacto de las desventajas sea menor o nulo cuando sea posible (2).

En la modalidad mixta (presencial/en línea) los estudiantes asisten al aula para tener sesiones de discusión guiadas por el profesor, en base al contenido del curso en la Web. Esto implica que los estudiantes pueden disponer de los contenidos de las páginas del sitio Web así como realizar las actividades y experiencias de aprendizaje programadas en él.

Los principales factores que intervienen en la enseñanza distribuida por Internet son la estructura del curso, las actividades de aprendizaje, los materiales didácticos que se ofrecen, las pautas y estrategias de comunicación, la evaluación y el rol del docente tutor y/o moderador. Esta nueva modalidad requiere también nuevas estrategias de enseñanza, centradas en el diseño de

materiales con las instrucciones necesarias para su buen manejo (18).

Trabajo Colaborativo

Según Marx, trabajo colaborativo son “múltiples individuos trabajando juntos de una manera planificada, en un mismo proceso o en procesos distintos pero conectados” (14). En base a esto, podemos afirmar que participar en un proceso colaborativo, persigue el desarrollo de conocimiento compartido (3). Cuando colaborar y compartir con otros lleva a considerar nuevas alternativas y reelaborar ideas tanto para comunicarlas como para fundamentar lo que se propone, es ahí que la interacción entre pares resulta genuinamente significativa y se torna relevante para resolver problemáticas. La mediación con los pares es el medio adecuado para desplegar el potencial individual y grupal, incrementar sus logros, la cohesión, y la solidaridad como practica social (19).

La sociedad modela su carácter conforme a cómo es la interacción entre las personas. (9). Considerando que son necesarias varios participantes para solucionar un problema complejo, es necesario que cada una de estas personas tenga un área de experticia diferente (20).

Denominamos, entonces, al trabajo colaborativo (*Groupware*) en términos de tareas comunes de un grupo y la necesidad de un entorno compartido. Se empieza a cambiar el enfoque, en donde se deja de pensar en la computación como herramienta para resolver problemas solamente, sino que, se empieza a ver también como herramienta para facilitar la interacción humana (9).

Aprendizaje Colaborativo

Los objetivos que persigue el aprendizaje colaborativo se centran en el desarrollo de estrategias de comprensión y explicación, de preguntas y respuestas. La discusión y el debate sirven para desarrollar las habilidades de comunicación, para desarrollar competencias intelectuales y profesionales y para promover el crecimiento personal del estudiante (13).

El aumento creciente de las tecnologías de las comunicaciones y su próspera integración con los entornos de trabajo colaborativo abren nuevas puertas a los medios de aprendizaje, permitiendo integrar muchas herramientas ya existentes con las crecientes tecnologías, brindando así mayor potencial a las herramientas de enseñanza. Con todo esto comienza a surgir la necesidad de brindar al alumno nuevas herramientas para permitir una interacción con sus pares o sus docentes y poder continuar con el proceso de aprendizaje sin dejar de lado la expansión en las TIC y los espacios colaborativos.

Herramientas CASE en Entornos Académicos Colaborativos

Considerando que la ingeniería de software requiere cada vez más participación grupal, conforme se complejizan los sistemas, se requiere de más gentes para resolver los problemas y esto implica mayor organización y gestión para lograr mejores resultados (20). A pesar de que se pueden obtener beneficios de cada tipo de herramienta individual, el mayor poder de las herramientas CASE está en la integración entre distintas herramientas (16). Las tendencias metodológicas de los últimos años están altamente influidas por el

enfoque de Orientación a Objetos (OO), esta visión permite identificar las abstracciones necesarias para proponer una solución consistente con los elementos del espacio del problema. (17). Toda herramienta CASE de modelado, debería soportar algún lenguaje de modelado estándar que permita simplificar tanto la comunicación como la documentación del sistema objeto de estudio. Muchas empresas dedicadas al desarrollo de software y distintos proveedores de herramientas CASE lo adoptaron y, por este motivo, se convirtió en un estándar mundial (12).

Existe un número significativo de trabajos que estudian la evolución de las herramientas CASE como soporte para la ingeniería de software, evolucionando también la colaboración como herramienta para mejorar los procesos de desarrollo, integrándolas en entornos WEB, con todas sus implicancias a nivel tecnología y ubicuidad. Las herramientas actuales de modelado carecen de utilidades destinadas a la enseñanza de la ingeniería de software (8).

La planificación y ejecución de procesos de enseñanza y aprendizaje para cursos de ingeniería de software, plantean un gran desafío a los docentes universitarios involucrados. La necesidad de una actualización dinámica de los contenidos no debe provocar el descuido de conceptos básicos vinculados a los principios fundamentales del desarrollo de sistemas de software (7). En cuanto a la aplicación de herramientas CASE como soporte de modelado para enseñar ingeniería de software, suelen utilizarse las herramientas presentes en el mercado (de uso profesional) o algunas herramientas de modelado que no son tan comunes en el mercado; por este motivo surge la necesidad de contar con un conjunto de herramientas CASE que permitan ayudar al alumno adquirir y

afianzar los conocimientos dentro de un espacio que permita tanto a sus pares como al docente colaborar para que se afiancen los conocimientos.

Actualmente existen plataformas de trabajo colaborativo y plataformas de modelado UML, se propone, por lo tanto, la integración de ambas dentro de un entorno académico mixto, en donde la herramienta no reemplace al modelo presencial sino que brinde un espacio de colaboración que permita optimizar tanto las relaciones entre docentes y alumnos, docentes y docentes, alumnos y alumnos, como el modelo de evaluación y seguimiento de trabajos.

Siendo UML una herramienta documental que tratar de reducir la ambigüedad comunicacional, y los aspectos colaborativos ayudan al desempeño tanto del docente como del alumno en los procesos de aprendizaje (5); se promueve la necesidad de integrar la tecnología UML dentro de una herramienta CASE, en el mismo entorno colaborativo, rescatando como características distintivas las ventajas de la ubicuidad, tiempo de respuesta, trazabilidad de evaluación, permita mejorar el uso de modelos.

Líneas de Investigación, Desarrollo e Innovación

El presente trabajo está radicado como proyecto de investigación y desarrollo en el CAETI (Centro de Altos Estudios en Tecnología Informática), de la UAI, dentro de la línea de investigación “Sociedad del Conocimiento y Tecnologías aplicadas a la educación”.

Resultados y Objetivos

Como objetivo general, el proyecto propone el desarrollo de un marco teórico que fundamente la bondades del

aprendizaje colaborativo en el contexto de la enseñanza de la ingeniería de software, además del desarrollo, diseño e implementación de una herramienta CASE colaborativa multiplataforma, denominada *UAICase*. Como objetivos específicos, nos planteamos: a) diseñar una plataforma académica colaborativa multiplataforma para evaluación, seguimiento, interacción y coordinación de proyectos informáticos, b) proponer una metodología de desarrollo de sistemas (no prescriptiva, sino académicamente orientativa) para ser utilizada en las asignaturas de cierre de cada ciclo (3° y 5° de la facultad de TI de UAI) y, por último, c) dar visibilidad al trabajo desarrollado por la facultad de TI de la UAI de modo tal que la herramienta CASE pueda ser compartida y utilizada por otras carreras de informática.

Formación de Recursos Humanos

El equipo de trabajo está formado, además de los integrantes que encabezan esta presentación, por los docentes de las asignaturas vinculadas con el proyecto de integración curricular. También participarán 10 alumnos de las diferentes asignaturas de la carrera. Además, el ingeniero Nicolas Battaglia está desarrollando su tesis de maestría en esta temática.

Referencias

1. Almenara, J. C. (1996). Nuevas tecnologías, comunicación y educación. Edutec. Revista electrónica de tecnología educativa
2. Aretio, L. G. (1990). Objetivos y funciones de la educación a distancia. In Filosofía de la educación hoy: entorno filosófico y contexto pedagógico (pp. 44-48). Universidad

- Nacional de Educación a Distancia, UNED.
3. Bannon, L. J., & Schmidt, K. (1989). CSCW-four characters in search of a context. *DAIMI Report Series*, 18(289).
 4. Barbera, E. (2006). Los fundamentos teóricos de la tutoría presencial y en línea: una perspectiva socio-constructivista. *Educación en red y tutoría en línea*, 161-180.7
 5. Castaño Mejía, J. A. (2008). Aldedo: herramienta colaborativa para modelado UML.
 6. Collazos, C. A., & Mendoza, J. (2006). Cómo aprovechar el “aprendizaje colaborativo” en el aula. *Educación y educadores*, 9(2), 61-76.
 7. Daniele, M., Uva, M., Martelloto, P., & Picco, G. (2010). Aplicación de herramientas CASE a la enseñanza de Ingeniería de Software: Gestión de la Configuración de Software y Testing Funcional. In V Congreso de Tecnología en Educación y Educación en Tecnología.
 8. Del Ben, E., & Vinjoy, M. (2011). Tecnología para la enseñanza de bases de datos: Un entorno de desarrollo enfocado a la enseñanza del diseño de bases de datos. In XIII Workshop de Investigadores en Ciencias de la Computación.
 9. Ellis, C. A., Gibbs, S. J., & Rein, G. (1991). Groupware: some issues and experiences. *Communications of the ACM*, 34(1), 39-58.
 10. Figueroa, S. G., & Cordero, R. D. (2012). Informática ubicua: su aplicación en el contexto universitario. In XIV Workshop de Investigadores en Ciencias de la Computación.
 11. Filippi, J. L., Lafuente, G. J., & Bertone, R. A. (2010). Diseño de un ambiente de aprendizaje colaborativo. In V Congreso de Tecnología en Educación y Educación en Tecnología
 12. Larman, C. (1999). *UML y Patrones. Introducción al análisis y diseño orientado a objetos*. Ed. Pearson
 13. Lavigne, G., Ovando, M. P. V., Sandoval, J. O., & Salas, L. M. (1970). Exploración preliminar del aprendizaje colaborativo dentro un entorno virtual/Preliminary study of collaborative learning in a virtual environment. *Revista Actualidades Investigativas en Educación*, 12(3).
 14. Marx, K (1849). *La Neue Rheinische Zeitung. Organ der Demokratie (Nueva Gaceta del Rin. Órgano de la Democracia)*," vol. 3, 4-26
 15. Oliveros, A., & Piccin, A. M. (2012). Web 2.0 para el ámbito académico: Implicancias en la Ingeniería de Software. In XIV Workshop de Investigadores en Ciencias de la Computación.
 16. Pressman, R. S. (2005). *Software engineering: a practitioner's approach*. Palgrave Macmillan.
 17. Quintero, J. B., de Páez, R. A., Marín, J. C., & López, A. B. (2012). Un estudio comparativo de herramientas para el modelado con UML. *revista universidad eafit*, 41(137), 60-76
 18. Rosales-Gracia, S., Gómez-López, V. M., Durán-Rodríguez, S., Salinas-Fregoso, M., &
 19. Rotstein, B., Scassa, A. M., Sáinz, C., & Simesen de Bielke, A. M. (2006). El trabajo colaborativo en entornos virtuales de aprendizaje. *Cognición*, 1(7), 38-45.
 20. Solano, A., & Collazos, C. (2013). Modelo para el diseño de actividades colaborativas desde un enfoque práctico. *Revista Universitaria RUTIC*, 1(2).

Plataforma de Gamificación para la Enseñanza de Programación en el Ámbito Universitario

Pablo M. Vera, Edgardo J. Moreno, Rocío A. Rodríguez,
M. Carina Vázquez, Federico E. Valles, M. Antonella Cornejo

GIDFIS (Grupo de Investigación, Desarrollo y Formación en Innovación
de Software)

Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza

Florencio Varela 1903, San Justo, Provincia de Buenos Aires
{pvera, emoreno, rrodriguez, mcvazquez, fvalles, mcornejo} @unlam.edu.ar

Resumen

La presente línea de investigación y desarrollo busca la aplicación de técnicas de gamificación a un curso universitario de programación con el fin de incentivar y motivar a los alumnos. Para ello se presenta el desarrollo de una plataforma web que mediante la corrección automática de ejercicios de codificación permite aplicar distintas técnicas como puntos, rankings, insignias y premios. De esta forma el alumno podrá ir siguiendo sus avances, comprobando si ha comprendido determinados temas y si es capaz de llevarlos a la práctica. Además las técnicas empleadas lo motivan a intentar ganar recompensas o mejorar su posición en los rankings siendo esto un estímulo para la utilización de la plataforma.

Palabras clave:

Gamificación, Programación, Educación

Contexto

El presente proyecto de Investigación y Desarrollo corresponde al grupo de

investigación GIDFIS (Grupo de Investigación, Desarrollo y Formación en Innovación de Software) perteneciente al Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza. El grupo es interdisciplinar al estar formado por docentes de distintos departamentos y en él se encuentran formándose en actividades de I+D alumnos de la universidad principalmente de Ingeniería.

Introducción

La enseñanza de programación requiere no solo un correcto dictado de los temas teóricos con ejemplificación práctica sino que además es necesario que el alumno se enfrente individualmente a los problemas e intente resolverlos con las herramientas aprendidas. Es por ello que es muy importante que el alumno realice prácticas con ejercicios propuestos por el docente para afianzar los temas dictados. El gran inconveniente que se presenta en hoy en día en la universidad es que muchas veces por falta de tiempo y otras directamente por desinterés, los alumnos se quedan solo con las prácticas realizadas en clase y únicamente estudian

en sus casas antes de los parciales. Una encuesta realizada a alumnos de primer año en las cátedras de programación durante el año 2015 sobre la cantidad de horas dedicadas al estudio de la materia comprueba esta afirmación. Donde el 45% de los alumnos no dedican tiempo adicional fuera de la clase para la resolución de ejercicios salvo antes de los parciales. En la tabla 1 pueden verse los resultados obtenidos en dicha encuesta.

Tabla 1. Resultados de la encuesta realizada a alumnos sobre las horas dedicadas a la materia fuera del aula.

Horas Dedicadas	Porcentaje
1h diaria	29,51%
Una semana antes del Examen	25,68%
2hs diarias	16,39%
Solo las horas de clases	13,11%
El día anterior al Examen	6,56%
3hs diarias	4,92%
4hs diarias	3,83%

Estos resultados ponen en evidencia que es necesario utilizar mecanismos que incentiven la participación de los alumnos motivándolos para mejorarse a sí mismos y entre sus pares. Es por eso que se decide estudiar como las técnicas de gamificación pueden aplicarse a un curso de programación universitario.

“La gamificación es la aplicación de mecánicas y dinámicas de juego en ámbitos que normalmente no son lúdicos” [1]. Es decir que muchas de las características presentes en los videos juegos se pueden aplicar a otros ámbitos para lograr una mayor participación, compromiso y motivación de las personas. Un ejemplo es el concepto de ganar puntos en un video juego donde se puede ir recolectando monedas o acumulando puntos por cumplir con

determinado objetivo y luego esos puntos sirven para mejorar un personaje, comprar accesorios o simplemente para estar mejor posicionado en un ranking. Este concepto de puntos es aplicado fuera de los videos juegos por muchas empresas para la fidelización de los clientes. Por ejemplo los bancos ofrecen puntos que se van sumando al realizar gastos con determinada tarjeta de crédito. Estos puntos luego pueden ser canjeados por premios o descuentos (otro concepto sacado de los videos juegos). De esta forma, el cliente elige este banco sobre la competencia e incentiva además la utilización de ese medio de pago sobre otros disponibles.

En el ámbito educativo estas técnicas pueden traer muchos beneficios. “Se puede lograr que las personas se involucren, motiven, concentren y se esfuercen en participar en actividades que antes se podrían clasificar de aburridas y que con la gamificación pueden convertirse en creativas e innovadoras” [2].

“La gamificación puede ser una estrategia de gran alcance que promueva la educación entre las personas y un cambio de comportamiento” [3].

Lo importante es que la gamificación puede producir cambios de hábito en los estudiantes, convirtiéndolo en un actor más activo en el proceso de enseñanza-aprendizaje [4], es decir motivarlo a participar.

Existen tres elementos de juegos que son necesarios para llevar a cabo lo mencionado: dinámicas, mecánicas y componentes.

Kevin Werbach [5], analista en tecnología, desarrolló la “Pirámide de Elementos de Gamificación” que podemos ver en la Figura 1.

Las dinámicas son el concepto, la estructura implícita, las mecánicas son los procesos que empujan el desarrollo y los

componentes son las implementaciones específicas de las dinámicas y mecánicas, que a veces dependiendo de los autores se considera como una mecánica en sí: avatares, insignias, puntos, colecciones, combates, rankings, niveles, equipos, bienes virtuales, son ejemplos de ello.

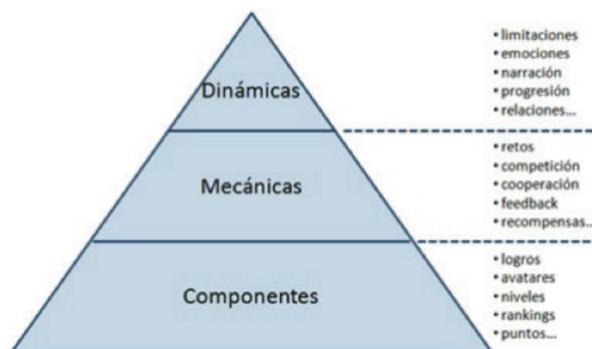


Figura 1. Pirámide de Elementos de Gamificación

Líneas de Investigación, Desarrollo e Innovación

Los ejes principales del trabajo son:

- Análisis de Técnicas de Gamificación Aplicadas a la Educación
- Generación de un Entorno Virtual que favorezca la implementación de las técnicas de gamificación.
- Desarrollo de Contenidos Temáticos Específicos para la plataforma
- Generación de niveles y desafíos
- Análisis de estrategias para premios y recompensas.

Resultados y Objetivos

A fin de aplicar las técnicas de gamificación a un curso de programación se decide desarrollar una plataforma de gamificación basada en codificación de funciones por parte de los alumnos. Para

lo cual se provee de un entorno web en el cual el alumno visualiza el enunciado del problema y además el prototipo de la función con su nombre y parámetros. Respetando esos datos deberá construir la función encomendada.

En base a los resultados obtenidos por el código desarrollado se otorga una serie de puntos al alumno lo que permitirá gamificar el curso utilizando distintos mecanismos como por ejemplo rankings.

Descripción general de las funcionalidades:

Los desafíos consisten en el desarrollo de una función determinada. Se dispone de un enunciado que explica el objetivo de la función a desarrollar y además se indica el tipo y nombre de los parámetros que la función recibe y el tipo de dato de retorno si corresponde. El alumno dispondrá de un área donde codificar la función. Una vez finalizada la codificación se envía el resultado y de forma automática se valida si la función fue correcta o no, dando un feedback al alumno de los posibles errores que cometió.

Esta funcionalidad se lleva a cabo utilizando el propio compilador de C que utilizan los alumnos para probar sus ejercicios. El procedimiento para crear un desafío es el siguiente:

1. El docente arma un enunciado de la función a desarrollar, especificando su prototipo incluyendo el nombre de los parámetros formales.
2. Para poder automatizar la prueba, el docente prepara un programa que incluye el prototipo de la función y la utiliza, pero no tiene el desarrollo de la misma.
3. El programa deberá enviar parámetros a la función y comprobar que el resultado sea correcto, armando así uno o más

casos de prueba. Si algún caso de prueba falla, entonces el programa creado por el docente retornará distintos códigos de error para poder identificar los mismos.

4. El docente carga mediante la plataforma el desafío, subiendo el archivo creado que incluye los casos de prueba.
5. Por cada código de error que puede retornar el programa, se realiza una descripción sencilla para que sea informado al alumno en el caso que ocurra.

Cuando el alumno envía el código desde la interfaz web, el código de dicha función es concatenado al código fuente subido por el docente que contiene los casos de prueba, de forma que la función sea utilizada. Ese archivo es entonces compilado automáticamente usando compilación por línea de comandos, en el caso que haya un error de compilación se muestra directamente dicho error al alumno. Si la compilación es exitosa, el programa es ejecutado y se analiza el código que retorna el programa. Si retorna 0 no hubo error, pero si retorna algún código de error detectado por los casos de prueba pre-establecidos, entonces se le muestra al alumno la descripción del error configurada por el docente.

Se definen una cantidad de ejercicios por cada tema de la materia (nivel). El docente podrá ir habilitando los distintos ejercicios a medida que avance la cursada. Con esta herramienta automatizada es posible aplicar distintas técnicas de gamificación, dado que la resolución de cada ejercicio brindará puntos al alumno.

Teniendo un puntaje máximo por cada ejercicio el mismo se irá decrementando según:

- Errores de compilación

- Errores lógicos detectados por los casos de prueba
- Errores por Time-Out
- Errores inesperados
- Tiempo de resolución superior al propuesto

Mecanismos de Gamificación a emplear:

Esta plataforma de corrección automática y asignación de puntos según el resultado obtenido, permite emplear distintas técnicas de gamificación. Entre ellas:

Rankings

- Los 10 puntajes más altos
- Los 10 puntajes más altos por nivel

Insignias

- Una insignia por cada nivel que el alumno supere (al completar todos los ejercicios de un tema dado)
- Una insignia por velocidad de resolución (al resolver un ejercicio correctamente en un tiempo inferior al propuesto)
- Una insignia por constancia, al entrar en forma periódica a la plataforma e iniciar 5 desafíos a lo sumo al día siguiente en que el docente lo habilita.
- Una insignia final al terminar con todos los temas de la materia

Herramientas Sociales

- Desafíos grupales (el alumno puede vincularse a un grupo y se mostraran puntos acumulados por cada grupo)
- Posibilidad de vinculación con redes sociales, publicando sus logros e insignias en facebook o twitter.

Premios

Al completar cada ejercicio (con un mínimo de puntos establecido) se le dará al alumno un premio al azar que

consistirá en tarjetas con las siguientes ventajas:

- Un minuto adicional para la resolución de un ejercicio
- Cinco minutos adicionales para la resolución de un ejercicio
- Puntos extra (esta tarjeta la podrá utilizar antes de enviar su código por primera vez y permitirá incrementar en un 10% los puntos obtenidos en el ejercicio que está realizando).

Actualmente esta plataforma se encuentra en desarrollo ya habiendo culminado la funcionalidad de corrección automática de ejercicios y asignación de puntos y trabajando actualmente sobre la plataforma de acceso para los alumnos y la aplicación de las distintas técnicas de gamificación mencionadas.

Formación de Recursos Humanos

El grupo está formado por 9 personas.

- 5 Docentes pertenecientes a diferentes departamentos conformando un grupo interdisciplinario.
- Una asesora externa experta en el área de gamificación perteneciente a la universidad de La Laguna, España.
- 3 alumnos de ingeniería quienes se están formando en actividades de Investigación y Desarrollo los cuales cuentan con becas asignadas

Vinculado con esta temática se encuentra en realización una tesis de maestría.

Referencias

1. Deterding, S., Khaled, R., Nacke, L. E., Dixon, D. (2011, May). Gamification: Toward a definition. In CHI 2011

Gamification Workshop Proceedings (pp. 12-15).

2. Werbach, K (2013), Gamificación. Fundació Factor Humà. Unidad de Conocimiento.

3. LEE, Joey J., et al. Greenify: A Real-World Action Game for Climate Change Education. Simulation & Gaming, 2013,

4. Cortizo Pérez, J. C., Carrero García, F. M., Monsalve Piqueras, B., Velasco Collado, A., Díaz del Dedo, L. I., & Pérez Martín, J. (2011). Gamificación y Docencia: Lo que la Universidad tiene que aprender de los Videojuegos.

<http://abacus.universidadeuropea.es/handle/11268/1750>

5. Werbach, K. (2013). Gamification

Pruebas y materiales instruccionales en ambientes virtuales de aprendizaje adaptativos

Huapaya Constanza⁽¹⁾, Benchoff Esther⁽¹⁾, Lizarralde Francisco⁽¹⁾, González Marcela⁽²⁾, Guccione Leonel⁽¹⁾, Lazurri Guillermo⁽¹⁾, Vivas Jorge⁽²⁾, Remón Cristian⁽¹⁾

⁽¹⁾Grupo de Investigación en Inteligencia Artificial Aplicada a Ingeniería /
Departamento de Matemática/ Facultad de Ingeniería

Juan B. Justo 4302, 0223 4816600 int 259

⁽²⁾Facultad de Psicología

Complejo Universitario - Funes 3250 - Cuerpo V, T.e. 0223) 475- 2526
Universidad Nacional de Mar del Plata

constanza.huapaya@gmail.com, ebenchoff.sead@gmail.com, francisco.lizarralde@gmail.com,
mpgonza@mdp.edu.ar, leonel.guccione@gmail.com, guillesky@gmail.com, jvivas53@gmail.com,
remoncristian@gmail.com

Resumen

Actualmente, las personas aprenden y trabajan en cualquier momento y en cualquier lugar usando dispositivos como computadoras de escritorios, laptops, teléfonos móviles, tabletas, etc. Estos dispositivos necesitan información sobre los usuarios a fin de satisfacer sus necesidades. Un modelo del usuario que reconozca las necesidades individuales de aprendizaje es la clave para la personalización en los sistemas educativos adaptativos.

Nuestra investigación propone el diseño y desarrollo de contenidos académicos personalizados, para estudiantes universitarios, a través de la plataforma Moodle. Para lograr este objetivo se ha definido un ciclo de desarrollo que aumenta el nivel de adaptación en cada iteración. La adaptación considera un motor (algoritmo) de adaptación cuyas entradas son los tipos de dispositivos que usan los estudiantes (pcs, laptops y teléfonos móviles), el estilo de aprendizaje, las preferencias y el diagnóstico cognitivo. La salida comprenderá los contenidos

didácticos personalizados (materiales y pruebas). Para este desarrollo se han contemplado nuevos enfoques del aprendizaje mediado (con TICs): Minería de datos educativa (MDE), Lógica e inferencia difusa, m-learning, Ambientes virtuales inmersivos 3D, representación difusa del dominio y adaptaciones de la plataforma Moodle.

Palabras clave: sistemas de aprendizaje adaptativos, lógica difusa, ambientes virtuales 3D, minería de datos educativa, m-learning, Moodle

Contexto

Nuestra línea de I/D/I se encuentra inmersa en el desarrollo de Ambientes Virtuales de Aprendizaje orientados a la enseñanza/aprendizaje de la ingeniería. En particular, el proyecto actual en el cual está incluida la investigación se denomina “Adaptación en un ambiente virtual de aprendizaje: pruebas y materiales personalizados“. La evaluación del proyecto se lleva a cabo en la Universidad Nacional de Mar del Plata.

Introducción

Cada estudiante es un individuo único con diversos intereses, experiencias, y logros. Los contenidos de estudio tradicionales, en general, no consideran estas diferencias. Actualmente se procura que los materiales instruccionales asistan, enriquezcan y extiendan el currículo del estudiante a fin de mejorar sus habilidades y conocimiento bajo la consideración de sus características individuales.

Adaptación y personalización en Ambientes Virtuales de Aprendizaje (AVAs)

Ser adaptativo es una propiedad que define la habilidad de cambiar ajustándose a diferentes condiciones. Esto es, algo es adaptativo si es capaz de cambiar para ajustarse, por sí mismo o no, a las circunstancias. En los ambientes de aprendizaje se puede distinguir dos tipos de ajustes. En primer lugar, hay *sistemas adaptables* que pueden cambiar sus parámetros cuando el usuario modifica al sistema de acuerdo a sus necesidades. En segundo lugar, los *sistemas adaptativos* implican adecuar automáticamente el sistema al usuario. Las necesidades del usuario son inferidas por el sistema, a través de la observación de su comportamiento, cuando este lo cambia de acuerdo a sus necesidades (Fröschl, 2005). En los sistemas e-learning se usan ambos conceptos. La adaptación en sistemas web AVA puede verse, primero, como la adaptación de los contenidos (la cual trata de seleccionar los contenidos más relevantes a las necesidades del estudiante), y en segundo lugar, la adaptación de la presentación de los contenidos (esto es, decidir cómo presentarlos de la forma más efectiva para el aprendizaje del alumno). Por otro lado,

la *personalización* es entendida como la adecuación de los contenidos o visualización del sistema a las preferencias del usuario. En particular, la personalización web trata con la forma de proveer los contenidos, esto es, que estilo y formato de su presentación.

El proceso de adaptación está basado en información almacenada en el modelo del usuario/estudiante y organizada en modelos específicos. Estos modelos dan la posibilidad de distinguir entre los estudiantes y proveer al sistema de la habilidad de adecuar su reacción a la actividad del usuario (Brusilovsky y Maybury 2002). En nuestro caso, para alcanzar un buen nivel de adaptación se analiza la actividad del estudiante a fin de detectar patrones regulares (o irregulares) sobre comportamientos individuales y grupales. Por ejemplo, se pueden identificar errores frecuentes, eficacia de estrategias tutoriales, granularidad de la representación del dominio, tópicos prioritarios, etc.

Actualmente nuestros desarrollos web pueden ser activados desde diversos dispositivos (computadoras, teléfonos móviles y tabletas).

Adaptación de los materiales y pruebas

El proceso de adaptación se inicia con una primera iteración definiendo la representación del dominio a enseñar en el modelo (FR-CN) (ver figura 1). A continuación se adquieren los estilos de aprendizaje con el cuestionario de Felder-Soloman (Soloman y Felder, 2003) y las preferencias individuales con cuestionarios específicos. Con esta información se construyen los primeros materiales con un grado bajo de adaptación. Ya es posible inicializar el motor de adaptabilidad y entregar los materiales a los alumnos. Posteriormente se adquieren los datos de la interacción

de los estudiantes con estos contenidos y los resultados se almacenan en las bases de datos de Moodle. Luego, con técnicas de MDE para filtrar resultados y con reglas de inferencia difusa, se computa el *diagnóstico cognitivo*. En particular, se usa KEEL (Knowledge Extraction based on Evolutionary Learning; <http://www.keel.es/>).

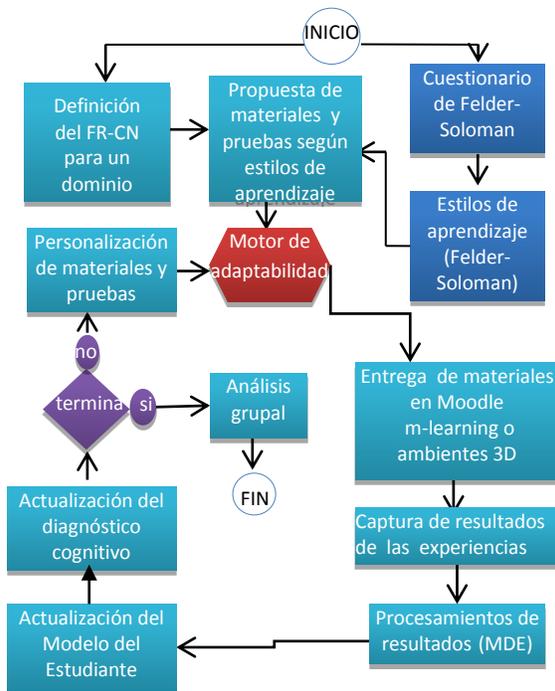


Fig. 1: ciclo de mejoramiento de la adaptabilidad

El diagnóstico cognitivo, simbolizado en un *perfil cognitivo*, da lugar a un incremento de la adaptación. Se ha considerado la presencia de la *incertidumbre* debido a que la valoración se basa en recortes del dominio, observaciones incompletas e inciertas, conjeturas e inferencias. Para comenzar el tratamiento de la incertidumbre estamos usando la Lógica Difusa en una red de conceptos que representa la estructura de los contenidos a enseñar y las dependencias entre los tópicos definidos. La Red de Conceptos Relacionados en forma Difusa (FR-CN,

Fuzzy Related-Concepts Network, Chrysafiadi y Virvou 2015) consiste de nodos que representan los conceptos del dominio de los materiales instruccionales, y arcos dirigidos que representan las relaciones entre conceptos de los contenidos de aprendizaje (ver fig.2). Las relaciones de dependencia permiten que el sistema tutorial identifique como el conocimiento de un concepto es afectado por nivel de conocimiento del estudiante o su relación con otros conceptos. Una relación de dependencia se caracteriza por el signo '+' o el signo '-' y un número (fuerza de impacto).

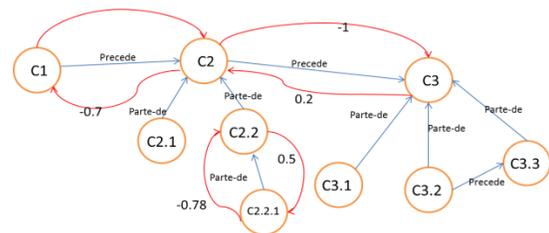


Fig. 2: modelo de representación del dominio y sus relaciones difusas (FR-CN)

En base a la red FR-CN se ha definido un modelo del estudiante (ver figura 3). El diagnóstico cognitivo de cada estudiante es estimado dinámicamente a partir de las evaluaciones de cada estudiante en cada nodo, generando el perfil cognitivo.

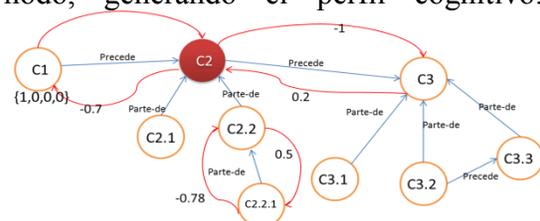


Fig. 3: modelo del estudiante

El mecanismo de cálculo en cada nodo se estima con la 4-upla = {desconocido, parcialmente conocido, conocido, aprendido} donde cada componente es un conjunto difuso. En la figura 3 el nodo C1 tiene la 4-upla = {1,0,0,0} indicando que el estudiante desconoce el concepto C1. Asimismo, hay nodos prioritarios por

su importancia en la enseñanza y evaluación. En la figura 3 se consigna que el nodo C2 debe ser evaluado.

En este punto, el motor de adaptabilidad (basado en un sistema de inferencia difusa, Stathacopoulou, 2007) ya posee sus tres variables lingüísticas de entrada: Estilos de aprendizajes, Diagnóstico cognitivo y Preferencias individuales. Se han definido variables lingüísticas de salida a fin de caracterizar los materiales y pruebas (p.e. perfil cognitivo, preferencias personales, nivel de dificultad, visualización, estructura, calidad, medio de entrega, etc.). Las salidas son materiales y pruebas personalizadas, las cuales son entregadas a los estudiantes. En las figuras 4 y 5 se ilustran dos casos desarrollados para alumnos de ingeniería.

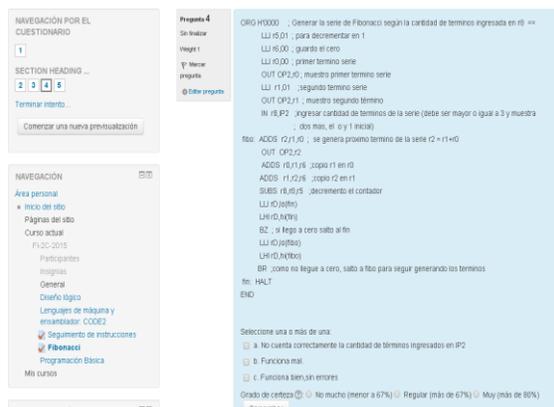


Fig. 4: prueba personalizada en Moodle

A partir de este punto se inician sucesivas iteraciones en pos de profundizar la adaptación de pruebas y materiales.

Líneas de Investigación, Desarrollo e Innovación

✓ **Adaptación en E-learning y M-learning:** se investiga el grado de profundidad de la adaptación de materiales (adaptación manual,

semiautomática o automática). Asimismo, se analiza y define indicadores de calidad para cada ejercicio o material según el Sistema de Indicadores para la Educación Superior (Proyecto INFOACES).

✓ **Ambientes virtuales inmersivos 3D:** En muchos sistemas de inmersión 3D se admite la creación de contenidos propios y la interacción multiusuario mediante texto, audio y video (Hew K. y Cheung W, 2010). Las simulaciones en 3D posibilitan la superación de las limitaciones espaciales y temporales del mundo real facilitando la experimentación y el desarrollo de un aprendizaje más significativo de diversos contenidos teóricos. Por ejemplo (ver figura 5), se ha desarrollado en nuestro grupo de investigación, una implementación no tradicional de una simulación del problema de los tres cuerpos, en el entorno virtual 3D Open Cobalt.

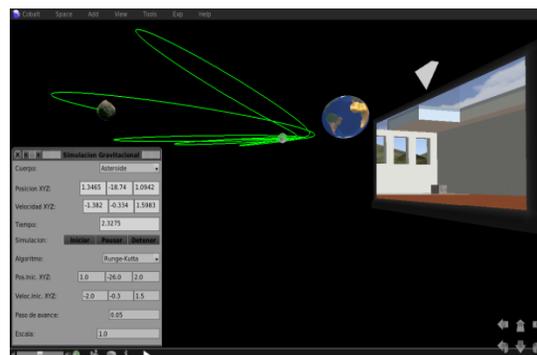


Fig. 5: espacio de la simulación gravitacional

✓ **Nuevos modelos de representación del dominio, el modelo del estudiante y diagnóstico cognitivo:** El proceso de estimar el diagnóstico se basa en observaciones inciertas, mediciones, conjeturas e inferencias. La presencia de la *incertidumbre* en el diagnóstico cognitivo se incrementa en un sistema educativo

adaptativo/personalizado debido tanto a la interacción indirecta entre los estudiantes y los profesores como a las dificultades técnicas (Grigoriadou et al. 2002). A fin de tratar la incertidumbre se usa la Lógica Difusa.

- ✓ **Minería de Datos Educativo (MDE):** Para procesar la vasta cantidad de datos obtenida por la actividad de los estudiantes en Moodle usamos Minería de Datos Educativo (MDE) (Romero, 2010) a fin de descubrir información valiosa “escondida” en las bases de datos de un AVA.

Resultados y Objetivos

- ✓ Desarrollar un modelo para optimizar mediante la adaptación la entrega de material instruccional personalizado en ambientes e-learning y m-learning.
- ✓ Definir y desarrollar un modelo de material de estudio y pruebas personalizado en temas de ingeniería y psicología.
- ✓ Desarrollar materiales especiales como sistemas de simulación educativa en Ambientes Virtuales 3D Colaborativos.
- ✓ Investigar la adaptación dinámica con mecanismos de transcoding y su aplicación.

Formación de Recursos Humanos

En el equipo de investigación hay especialistas en educación, informática, ingeniería y psicólogos cognitivos. Asimismo hay un estudiante de la carrera de Ingeniería Informática integrado al grupo de investigación quien colabora activamente

Actualmente hay tres postgrados cursándose en la Facultad de Informática de la Universidad Nacional de La Plata:

dos Magísteres en Tecnología Informática Aplicada en Educación a punto de finalizar y un Magister en Ingeniería de Software en curso.

Referencias

Brusilovsky P. y Maybury M.T. (2002) From adaptive hypermedia to the adaptive web. *Communications of the ACM*, vol. 45, no. 5 p.p. 30–33.

Chrysafiadi K.; Virvou M., (2015). *Advances in Personalized Web-Based Education*. Springer Cham Heidelberg.

Froschl C. (2005). User Modeling and User Profiling in Adaptive E-learning Systems. Master's Thesis de la Graz University of Technology.

Grigoriadou M, Kornilakis H., Kyparisia A. Papanikolaou K. y Magoulas G.(2002). *Fuzzy Inference for Student Diagnosis in Adaptive Educational Hypermedia*. Lecture Notes in Computer Science Volume 2308, 2002, pp 191-202. Methods and applications of artificial intelligence.

Hew K.F. y Cheung W. S.(2010) Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British Journal of Educational Technology Volume 41*, Issue 1, pages 33–55.

Soloman, B. A., y Felder, R. M. (2003). Index of learning styles questionnaire. Recuperado junio, 2015, desde <http://www.engr.ncsu.edu/learningstyles/ilsweb.html>.

Romero C. y Ventura S. (2010). Educational Data Mining: A Review of the State of the Art. *Ieee transactions on systems, man, and cybernetics—part c: applications and reviews*, vol. 40, no. 6, november.

Stathacopoulou R., Grigoriadou M., Samarakou M. y D. Mitropoulos. (2007). Monitoring students' actions and using teachers' expertise in implementing and evaluating the neural network-based fuzzy diagnostic model. *Expert System with Applications: 32(4)* 955-975

ENTORNOS VIRTUALES EN LA ENSEÑANZA DE LAS CIENCIAS EXACTAS Y NATURALES

Escudero, Silvia Susana; Marazzo, José Luis; Pompei, Sabrina;
Peri, Jorge Alberto

UNIVERSIDAD NACIONAL de LUJÁN, Dpto. de CIENCIAS BÁSICAS

Ruta 5 y Avenida Constitución - (6700) Luján, Buenos Aires, Argentina.

Teléfonos: +54 (02323) 423979/423171

sescudero@unlu.edu.ar jmarazzo@unlu.edu.ar sabrinapompei@gmail.com

jp28580@yahoo.com

Resumen

Se presentan los avances logrados en el año 2015 en la línea de investigación de las TICs en la enseñanza de las Ciencias Exactas enfatizando la medición del impacto en el aprendizaje medidos por los laboratorios virtuales en un grupo de estudiantes de escuela media pertenecientes a la generación de estudiantes denominada “Nativos Digitales” o “Inmigrantes Digitales” según el autor que se trate. También propiciar el trabajo colaborativo entre los docentes a través del foro del EVA. Por último se analizan los aportes de la programación lógica para el diseño a futuro laboratorios virtuales por alumnos de de la carrera Lic. en Sistemas de Información de la Universidad Nacional de Luján.

Palabras clave

TICs, Enseñanza de la Ciencias, Laboratorios Virtuales.

Contexto

Este trabajo forma parte del Proyecto de Investigación *Las TICs en la Enseñanza de las Ciencias Exactas* radicado en la Universidad Nacional de Luján. Se muestran avances del mismo y las líneas de investigación futuras.

Introducción

La evolución tecnológica ha generado importantes repercusiones en todas las áreas de la sociedad, y la educación no está ajena a este cambio. La mayoría de los estudiantes de nivel medio o superior pertenecen, desde el punto de vista digital, a una generación que es denominada según Prensky “**Nativos Digitales**”.

Para **Prensky** estos estudiantes representan la primera generación que creció con las nuevas tecnología y como resultado de ello no piensan ni procesan la información de la misma manera que sus predecesores, ya que sus patrones de pensamiento han cambiado. Frente a estos «nuevos» estudiantes **Nativos Digitales**, hablantes nativos de un lenguaje digital, encontramos a los **Inmigrantes Digitales**, personas que aunque puedan llegar a adaptarse y aprender a usar estas tecnologías, no dejan de ser inmigrantes en un mundo digital, manteniendo su «acento» que les diferencia.

Tapscott en cambio utiliza el término **Generación Net** para denominar a esa generación nacida a principios de los años 80, y por tanto «bañada en bits», caracterizados especialmente por su curiosidad, su capacidad de adaptación y su especial interés por los contextos tecnológicos. Describe, además, en su obra diez importantes características de esta generación, como por ejemplo la independencia y autonomía, la curiosidad e investigación, la suspicacia frente a los intereses corporativos, la

firmeza y definición de sus opiniones, la necesidad de inmediatez, entre otras cosas.

En definitiva, y más allá de definiciones, nuestros estudiantes conocen herramientas TICs y las saben utilizarlas. Por eso creemos que es necesario modificar las prácticas educativas en las aulas. *“Son las TIC, desde su concepción, diseño y posterior empleo en los procesos de aprendizaje, las que nos ayudan a adecuar la enseñanza a los nuevos escenarios de educación que están apareciendo”*.

Laboratorios Virtuales

Un *Laboratorio Virtual* es un simulador interactivo de un laboratorio real, donde los alumnos mediante tecnología web reciben información y realizan actividades interactivas de Física, Química, Matemáticas y Cs. Naturales en general.

Los programas de *Laboratorios Virtuales* permiten crear un enfoque constructivista del proceso enseñanza – aprendizaje donde el alumno puede constatar la hipótesis a través de experiencias virtuales.

Cabero señala una serie de posibilidades y ventajas que ofrecen los *Laboratorios Virtuales* las cuales podemos resumir en:

- a) La habilidad que los estudiantes inicialmente cuentan en el manejo de simuladores y herramientas informáticas les permite desenvolverse fácilmente en entornos tecnológicos.
- b) La predisposición que los alumnos muestran hacia el uso de las TICs.
- c) Posibilidad de realizar trabajos individuales y grupales entre los alumnos.

- d) Posibilidad de acceder a la realización de experiencias que de otro modo sería inaccesibles
- e) Repetición irrestricta de la experiencia hasta comprender el concepto.

Podemos decir que facilitan la realización de trabajos de laboratorio cambiando la imagen negativa que suelen tener los alumnos sobre los mismos.

El aporte de la Programación Lógica (PL)

En los cursos de las carreras de informática, el paradigma lógico de programación puede hacer un aporte interesante a esta problemática, debido particularmente a:

El lenguaje Prolog está diseñado para poder emular el razonamiento lógico.

Los alumnos de la carrera Lic. en Sistemas de Información de la Universidad Nacional de Luján pueden participar activamente en el diseño y construcción de laboratorios virtuales, lo que facilita la fecundación cruzada entre la enseñanza de la PL por una parte, y la utilización de la misma en otras áreas de la educación por la otra.

En el grupo de trabajo se trabajó en el desarrollo de interfaces amigables de Prolog (Edulog), que permitan su uso por personas con poca o ninguna experiencia en programación. (Godoy, 1999)

Los resultados obtenidos muestran que la PL resulta apta para modelar situaciones en un amplio espectro del conocimiento, desde al álgebra de relaciones y funciones hasta la resolución de acertijos, incluyendo temas de historia, biología y electrónica entre otros. (Peri, 2000)

Líneas de Investigación, Desarrollo e Innovación

Utilizando el EVA **e-educativa** licenciado por la Universidad Nacional de Luján se diseñó un repositorio de programas de laboratorios virtuales en línea y para aplicaciones móviles con una breve descripción de los mismos agrupados por área (Física, Química, Matemática y Ciencias Naturales) y dentro de cada uno de ellos por temas. (WICC 2015).

Para la selección de los recursos se tuvieron en cuenta los siguientes criterios:

a) Aspectos Tecnológicos:

- *hardware y software*, deben ser los mínimos requeridos para poder ser utilizados en cualquier equipo.
- *tiempo de acceso a los programas*: si la carga del programa lleva demasiado tiempo se pierde el interés del usuario.
- *Calidad de las imágenes, audio, animaciones.*
- *Transición entre las distintas pantallas*
- *Tamaño de la letra de los textos*
- *Su utilización requiere mínimos conocimientos previos.*
- *Fácil intervención de los usuarios*

b) Aspectos Pedagógicos

- *Contenidos adecuados al currículo oficial*
 - *La forma de presentación de los contenidos deben ser motivadores para los alumnos*
 - *Desarrollo de la creatividad*
 - *Nivel de dificultad adecuado a los conocimientos previos de los alumnos*

- *Metodología innovadora de presentación de los contenidos.*

Este espacio al que hemos llamado LABORATORIOS VIRTUALES es de acceso libre por lo cual toda persona interesada puede ingresar a explorar el mismo.

En esta segunda etapa del proyecto los objetivos son:

- 1) Difusión del espacio entre los docentes de nivel medio y universitario.
- 2) Propiciar el trabajo colaborativo entre los docentes a través del Foro del EVA donde ellos podrán opinar, presentar sus inquietudes y sugerencias el mismo estará coordinado por los integrantes del equipo de investigación.
- 3) Medir el impacto en el aprendizaje de los laboratorios virtuales en un grupo de estudiantes de escuela media. Las variables a tener en cuenta serán actitud hacia la ciencia, habilidad para resolver problemas, comprensión del tema.

Resultados

La producción se plasmará en publicaciones en congresos y eventos, y una tesis de Magister.

Al finalizar el proyecto se pretende:

- a) Disponer de una evaluación del uso de los laboratorios virtuales
- b) Desarrollo de una interface Edulog para Android, del cual ya se dispone de una primer versión experimental.

Formación de Recursos Humanos

El equipo de trabajo está integrado por un director, investigador categorizado en el programa de Incentivos del Ministerio de Educación, además docentes del Departamento de Ciencias Básicas con formación en Informática y Ciencias Exactas y estudiantes en proceso de formación.

El desarrollo de las versiones de Edulog, así como la exploración de las posibles aplicaciones, se hace en el ámbito del dictado de asignaturas de Programación Funcional y Lógica de dos universidades nacionales.

Referencias

Prensky, M. (2001). Digital natives, digital immigrants. En *The Horizon*,.

Tapscott, D. (1998). *Growing up digital: The rise of the net generation*. New York: McGraw-Hill.

Mondéjar, J., Mondéjar, J. A. & Vargas, M. (2006). *Implantación de la metodología elearning en la docencia universitaria: una experiencia a través del proyecto Campus Virtual*. Revista Latinoamericana de Tecnología Educativa.

Cabero, J. (2008) *Las TICs en la enseñanza de la química: aportaciones desde la Tecnología Educativa*. En Bodalo, A. y otros (eds.) (2007): *Química: vida y progreso Murcia*, Asociación de Químicos de Murcia.

Cabero, J. y Romero, R. (2005) *Criterios generales para el diseño, la producción y la utilización de las TICs en la enseñanza en Curso: TICS para la formación. Su utilización didáctica*. Universidad de Sevilla.

Coll, C., Mauri, T. y Onrubia, J. (2008): *El análisis de los procesos de enseñanza y aprendizaje mediados por las TIC: una perspectiva constructivista en Barberà, E., Mauri, T. y Onrubia, J. (coords.): Cómo valorar la calidad de la enseñanza basada en las TIC*, Editorial Grao, Barcelona.

Coll, C., Colomina, R., Onrubia, J. y Rochera, M.J. (1995). *Actividad conjunta y habla: una aproximación al estudio de los mecanismos de influencia educativa*. En P. Fernández Berrocal y Ma. Angeles Melero (Comps.), *La interacción social en contextos educativos*. Madrid: Siglo XXI.

Colomina, R., Onrubia, J. y Rochera, M. J. (2001). *Interactividad, mecanismos de influencia educativa y construcción del conocimiento en el aula*. En C. Coll, J. Palacios y A. Marchesi (comp.), *Desarrollo psicológico y educación 2. Psicología de la educación escolar* (pp. 437-458). Madrid: Alianza.

Godoy, D. (1999): *El intérprete Edulog*. Tesis de grado. Universidad Nacional de Luján.

Peri, J. A.; Godoy, D. (2000): *Utilización de acertijos lógicos como ejercicios motivadores para la enseñanza de la programación lógica*. CACIC.

Peri, J.; Escudero, S. ; Marazzo, J. L.; Pompei, S. (2015) *“Las TICs en la enseñanza de las Ciencias Exactas y Naturales”*. WICC.

Proyecto: Inteligencia Artificial y desarrollo de Simuladores hacia el diseño de Cursos Abiertos On Line

Laura C. Díaz, Adriana Chautemps

Facultad de Ciencias Exactas, Físicas y Naturales
Universidad Nacional de Córdoba
lcd_ic@yahoo.com.ar, achautemps@gmail.com,

Resumen

Este proyecto contribuirá a dar luz a la factibilidad de implementar Cursos Abiertos Masivos en línea (del inglés, MOOC), desde esta Unidad Académica. Esta modalidad de Educación impacta sobre: la visibilidad de la institución, la mejora en la accesibilidad por parte de ciudadanos del mundo entero, la oportunidad de formación continua a lo largo de la vida y la construcción colectiva y colaborativa de conocimiento. Además de poner accesible un conocimiento más acabado del estudiante, útil no solo para el diseño de los MOOC sino también para la gestión de la educación superior y para las prácticas cotidianas de aprendizaje y evaluación en sus diversas modalidades.

Por otra parte, las actividades en el ámbito de la Inteligencia Artificial impactarán como un valor agregado de capital humano a los estudiantes de grado que integran este equipo de investigación.

Por último, ofrecerá el desarrollo y la accesibilidad de una herramienta, íntegramente desde el ámbito universitario. El simulador para prácticas de laboratorio de energía nuclear destinado a la formación de tecnólogos para los cursos con modalidad presencial

y en la accesibilidad a estos cursos que quedarán disponibles en plataformas abiertas on line.

Las experiencias previas justifican y dan fundamentación a sus objetivos.

Palabras clave: Inteligencia Artificial, Simuladores, Cursos Abiertos.

Contexto

Las acciones previstas para el bienio 2016 - 2017 están concebidas como trabajo interdisciplinario entre los proyectos que conforman el Programa: Apropiación del Conocimiento y la Tecnología, en el seno de la Universidad Nacional de Córdoba (UNC), del que este proyecto forma parte. Las acciones de investigación están dirigidas, por un lado, a asignaturas masivas de diversas Unidades Académicas de la UNC, en las que están involucrados los integrantes del referido Programa. Por otra parte, desde el Centro Universitario de Tecnología Nuclear (CUTEN), se llevan a cabo las actividades de desarrollo tecnológico y formación en el ámbito de prácticas de Laboratorio en Energía Nuclear.

El proyecto está en proceso de evaluación para la acreditación y financiamiento en la convocatoria 2016 -

2017 de la Secretaría de Ciencia y Técnica (SECyT) de la UNC.

Introducción

El uso de Tecnologías de Información y Comunicación (TIC) en educación y el desarrollo de herramientas educativas constituye un ámbito de trabajo actualmente en continua evolución y de relevancia en la construcción de saberes. (Zabalza, 2013)

Las mejoras que generan las estrategias centradas en la incorporación de estas tecnologías remiten a aspectos asociados a la apropiación del conocimiento y de la tecnología por parte de la sociedad y a su construcción colaborativa. Emergen nuevos objetos de análisis y concepciones antropológicas que se construyen en esas sociedades. (Colobrans, 2011)

Su aplicación en educación para atender la masividad se contextualiza en un momento de construcción colectiva de saberes inmersos en la realidad de la región, en las heterogeneidades de la población estudiantil, en las formas narrativas y de comunicación; en síntesis, en la configuración de una nueva matriz asociada a la praxis de la apropiación del conocimiento y de la tecnología. En nuestro país, a través de la institucionalización del Programa de Internacionalización de la Educación Superior y Cooperación Internacional de la Secretaría de Políticas Universitarias del Ministerio de Educación, se orientan acciones en tal dirección.

En particular, con el surgimiento de los MOOC es posible pensar que ciudadanos de un continente cualquiera puedan tener acceso desde sus lugares de residencia a un nivel de acreditación de postgrado de universidades con prestigio internacional.

En este contexto, abocarse a indagar alrededor del estudiante como principal

actor del complejo escenario, depositario de estos saberes que se transmiten de generación en generación, atendiendo a su compromiso social y político, es parte de una de las líneas de investigación propuestas.

Un mejor conocimiento del estudiante a partir de los resultados del tratamiento de bases de datos con Tecnologías Inteligentes de Explotación de la Información (TIEI), implica un gran aporte para el desarrollo de los MOOC ya que posibilita adecuarlos a sus potenciales usuarios. En estas modalidades el único acercamiento entre el conocimiento y el estudiante por apropiarse de él, es el entorno virtual, que se constituye en vínculo de comunicación durante los procesos de aprendizaje y de evaluación.

Los resultados obtenidos en el proyecto antecesor de éste, “Inteligencia Computacional y TIC: Estrategias para Facilitar el Aprendizaje en la Universidad Nacional de Córdoba” alientan a profundizar y extender la investigación (Díaz, Martins, García Martínez, 2015), atendiendo ahora a las necesidades de información que manifiesten investigadores, docentes y estudiantes y particularmente las que favorezcan al diseño de los MOOC desde la UNC en campos de aplicación disponibles en el programa del que este proyecto es parte, a saber: Matemática en Ingeniería, Relación Público- Privada para el Desarrollo Tecnológico y Prácticas de Laboratorio en Energía Nuclear, constituyéndose en horizonte de estas acciones.

Otro tema de relevancia para la apropiación tecnológica es la mejora en la formación de los expertos, atento a satisfacer la demanda de especialistas en el país. El desarrollo e incorporación de simuladores a las clases presenciales y a distancia en el ámbito de los Laboratorios de Enseñanza de Energía Nuclear, contribuyen a fortalecer el aprendizaje en

esos espacios. Este tipo de tecnología es fundamental para adquirir destreza en prácticas de laboratorio.

La iniciativa surge también por la necesidad de capacitar a grupos numerosos, lo que dificulta el entrenamiento con instrumentos reales. Con el simulador el alumno puede repetir la experimentación hasta obtener la experticia necesaria que la situación requiera (Pérez Lucero, Chautemps, Bertone, Díaz, 2015).

Estas herramientas educativas contribuyen a la adquisición de conocimiento científico en la relación entre el docente, el material educativo y el estudiante (Ausubel, Novak, Hanesian, 1983). El sentido de esta relación es compartir significados entre el alumno y el profesor, respecto de los conocimientos vehiculados por los materiales educativos.

Este equipo de trabajo evidencia su compromiso con la formación en espacios virtuales abiertos, consciente de la necesidad de ofrecer modalidades virtuales, a través de la implementación de cursos en plataforma Moodle, en la Red Latinoamericana de Educación a Distancia (LANENT) y en el CUTEN.

Indagando antecedentes y relatando la propia experiencia:

Uno de los fenómenos relevantes, emblemático en el sentido de la accesibilidad al conocimiento, es la creciente oferta de cursos masivos que propicia entornos de aprendizaje abiertos para que miles de personas de todo el mundo sigan diferentes iniciativas educativas. Si bien nuestra Universidad participa en consorcios y redes (Díaz, 2014) sus avances son incipientes.

En relación al conocimiento del estudiante son relevantes los estudios realizados en otras universidades del país para la identificación de las características

socioeconómicas y demográficas más significativas en la construcción de su modelo de comportamiento académico, resultando información útil para el diseño de políticas públicas en educación superior. (Kuna, García Martínez, Villatoro, 2010).

En tal sentido, se detectaron relaciones de interés entre las características de los estudiantes de Ingeniería y de Ciencias Económicas de la UNC considerando dimensiones socioeconómicas, académicas, mediante la aplicación de Procesos de TIEI sobre las bases de datos de SIU- Guarani. Ambos casos de estudio en asignaturas masivas (Díaz, Las Heras, Bartó, 2015).

Identificar las capacidades de los estudiantes como factor significativo en la mejora del aprendizaje, hallazgos realizados en (Dehnadi y Bornat, 2006), despierta el interés en el uso de Redes Neuronales para el descubrimiento de patrones de comportamiento en la detección temprana de dificultades cognitivas del estudiante.

Los conocimientos que surgen de estas experiencias motivan a profundizar el uso de herramientas de Inteligencia Artificial como aspecto necesario para transitar el camino hacia el diseño de MOOC.

Por otra parte, el nuestro en particular en el ámbito de la Energía Nuclear, las herramientas educativas en simulaciones promueven el aprendizaje experimental ya que los contenidos conceptuales pueden ser comprendidos con más profundidad, a la vez que se aprenden contenidos procedimentales (Njoo y Jong, 1993).

La simulación se adapta a una estrategia de tipo exploratoria ya que oculta un modelo que debe descubrir el alumno, además de situarlo en un mundo intermedio entre lo concreto y lo abstracto que lo ayuda a progresar de la

fase concreta de su desarrollo a la fase formal (Valente y Neto, 1992).

Con los simuladores en Laboratorios de Energía Nuclear se cuenta con la posibilidad de repetir las condiciones de la realidad para adquirir competencias en ese ámbito en condiciones que evite, con la protección del riesgo inherente a la manipulación en este campo (Murúa, Chautemps y Odetto, 2013).

Líneas de Investigación, Desarrollo e Innovación

Las líneas de acción que abarca este proyecto se abocan a: 1° Avanzar hacia el diseño de los MOOC desde la UNC, como una estrategia para contribuir a la apropiación del conocimiento y la tecnología. 2° Aplicar Inteligencia Artificial para profundizar en el conocimiento del Estudiante, destinatario de esos cursos. 3° Desarrollar una herramienta educativa simulada, de aplicación en aulas virtuales o reales que mejore la oportunidad de aprendizaje en prácticas de laboratorio de Energía Nuclear.

Resultados y Objetivos

El objetivo general de este proyecto es facilitar la apropiación del conocimiento y la tecnología en Educación Superior, a través de acciones hacia la implementación de los MOOC.

Los Objetivos específicos consisten en:

- a) Profundizar en el descubrimiento del modelo del estudiante a partir de la situación socioeconómica, académica y cognitiva, usando Tecnologías Inteligentes, atendiendo a los requerimientos que manifiesten los expertos en el dominio.
- b) Contar con un diagnóstico que describa las dificultades, fortalezas, amenazas y oportunidades para

generar los MOOC en dimensiones: Reglamentaria, Institucional, de Recursos Físicos y de Recursos Humanos. En esta dirección, uno de los integrantes ha realizado acciones de investigación exploratoria en el ámbito de su carrera de Doctorado radicado en la Universidad de Lanús: Gestión de la Educación Superior en Contextos de Masividad Basada en Tecnologías Inteligentes de Transformación de Información.

- c) Atendiendo a los resultados del diagnóstico realizado, lograr avances hacia la construcción de cursos abiertos on line en los campos de aplicación de los seis proyectos que conforman el Programa Apropiación del Conocimiento y la Tecnología.
- d) A través de las acciones de diseño, desarrollo y validación, lograr un Simulador para el trabajo con instrumental de laboratorio. En este sentido se vienen desarrollando instrumentos simulados los cuales se implementan en aula durante clases presenciales.
- e) Facilitar el acceso a prácticas de laboratorio en cursos a distancia a fin de propiciar el conocimiento científico teórico-práctico en esos temas específicos. Hasta el momento se utilizan videos ilustrativos de las prácticas de laboratorio utilizando tanto instrumentos reales como simulados.
- f) Favorecer la formación de tecnólogos en las áreas de esta propuesta de investigación, focalizándose en los estudiantes que forman parte de este proyecto. Para ello se realizan capacitaciones en las diferentes áreas del desarrollo en tecnología educativa. Actualmente se está trabajando en una tesis de maestría sobre Tecnología Informática Aplicada en Educación.

Formación de Recursos Humanos

Este proyecto de Investigación está constituido en total por trece integrantes, ocho docentes y cuatro estudiantes de grado de la UNC y un docente de la Universidad Tecnológica Nacional Regional Villa. Se organizan en equipos en función de las líneas de investigación presentadas y los objetivos expuestos.

Actualmente están en desarrollo dos tesis de posgrado, una de Doctorado y una de Maestría y tres tesinas de grado.

Por otra parte, uno de los objetivos del proyecto es contribuir a la formación continua de los docentes, investigadores y estudiantes de nivel superior, a través de actividades extracurriculares para la transferencia de los avances y hallazgos resultantes, a ejecutar durante este bienio.

Referencias

- Ausubel, D., Novak, J. D., Hanesian, H. (1983): *Psicología educativa, un punto de vista cognoscitivo*. México: Ed. Trillas.
- Colobrans, Jordi (2011): *Viatges al Món d'Arduino. Què podrí aprendre de les cultures informàtiques?* Citilab de Cornellà. Cornellà de Llobregat, Barcelona. Etnografia.
- Dehnadi, Saeed y Bornat, Richard. (2006): *The camel has two humps (working title)*. School of Computing. Middlesex University, UK.
- Díaz L. (2014): *Una vuelteita más, no me quiero bajar. TIC y democratización de la Educación Superior*. VIII Congreso Iberoamericano de Docencia Universitaria y Nivel Superior. U.N.L.
- Díaz L., Las Heras J., Bartó C., (2015): *Inteligencia Artificial y TIC: Estrategias para facilitar el aprendizaje en la UNC*. II Jornadas nacionales experiencias e investigación en educación a distancia y tecnología educativa. UNC.
- Díaz L., Martins S., García Martínez R. (2015): *Descubrimiento de Patrones Socio-Económicos de Población Estudiantil de Carreras de Ingeniería Basado En Tecnologías de Explotación de Información*. En Memorias de Congreso TE&ET 2015.
- Kuna H, García Martínez R, Villatoro R, (2010): *Identificación de causales de abandono de estudios universitarios. Uso de procesos de explotación de información*. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología, TE&ET. Art. 5.
- Murúa, C., Chautemps, A. y Odetto, J. (2013): *Estrategia para la enseñanza semipresencial de física de reactores y radioprotección*. Memorias del Seminario Internacional RUEDA". Mendoza.
- Njoo, M. y Jong, T. (1993): *Exploratory learning with a computer simulation for control theory: Learning processes and instructional support*. Journal of Research in Science Teaching, 30, 821-844.
- Pérez Lucero, A., Chautemps, A., Bertone, R., Díaz, L., (2015): *Simuladores Aplicados en Laboratorios de Energía Nuclear*, Memorias del XXI Congreso Argentino de Ciencias de la Computación. Junín. Buenos Aires.
- Valente, M. y Neto, A.J. (1992): *El ordenador y su contribución a la superación de las dificultades del aprendizaje en mecánica*. Enseñanza de las ciencias, vol.10, n.1, 80-85.
- Zabalza M., (2013): VIII Congreso Iberoamericano de Docencia Universitaria y Nivel Superior. <http://www.iberoamericano2014.unr.edu.ar/>

Estudio de la influencia del uso de simulación en la enseñanza de redes de computadoras en el nivel universitario (resultados parciales)

Daniel Arias Figueroa¹, Javier Díaz², Cecilia Gramajo¹, Gustavo Gil¹, Ernesto Sánchez¹, Loraine Gimson¹, Álvaro Gamarra¹, Jorge Silvera¹, Diego Gil¹, Nelson Rodríguez³, María Murazzo³, Marcelo Moreno³, Miguel Guevara³

C.I.D.I.A. (Centro de Investigación y Desarrollo de Informática Aplicada)¹
Facultad de Ciencias Exactas de la Universidad Nacional de Salta

L.I.N.T.I. (Laboratorio en Nuevas Tecnologías Informáticas)²
Facultad de Informática de la Universidad Nacional de La Plata

Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.³
Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de San Juan

daaf@cidia.unsa.edu.ar, nelson@iinfo.unsj.edu.ar

Resumen

Una de las preocupaciones constantes de las universidades es estar a la vanguardia en los métodos de enseñanza y ofrecer oportunidades de prácticas innovadoras que apoyen la preparación de sus estudiantes para su futura vida laboral. Uno de los objetivos claves en la enseñanza de redes de computadoras en la carrera de grado de la Licenciatura en Análisis de Sistemas en la Universidad Nacional de Salta, es la de transmitir fundamentos y conceptos básicos a sus estudiantes (en contraste con una formación más específica de técnico en redes). Sin embargo, desde hace un tiempo venimos investigando cómo facilitar la relación entre la realidad y las teorías y modelos, es decir, entre lo concreto y lo abstracto.

Dentro de nuestro trabajo de investigación pretendemos evaluar la influencia de la utilización de software de simulación, en la enseñanza de contenidos de redes de computadoras en nuestra carrera de grado y cursos de postgrado. Para ello nos hemos planteado un estudio aplicado, longitudinal y experimental que, asumiendo la complejidad del fenómeno educativo, considere aspectos cuantitativos y cualitativos.

Palabras clave: protocolo tc-pip, simulación, virtualización, enseñanza, redes.

Contexto

La línea de investigación se encuentra apoyada por el C.I.D.I.A. (Centro de Investigación y Desarrollo de Informática Aplicada) que depende de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta y el L.I.N.T.I. (Laboratorio en Nuevas Tecnologías Informáticas) – Facultad de Informática de la Universidad Nacional de La Plata, y por lo tanto se cuenta con toda la infraestructura disponible para esta investigación. El proyecto se encuentra acreditado como Proyecto Tipo “A” N° 2230/0 en el CIUNSa – Consejo de Investigación de la Universidad Nacional de Salta y cuenta con su financiamiento.

Introducción

El rápido y continuo desarrollo tecnológico que las organizaciones viven, particularmente en las áreas de tecnología, altera la manera de enseñar y, por supuesto, de aprender.

En muchas universidades, la computadora, las redes, los sistemas de videoconferencias, la Internet y las soluciones de software se usan a diario. Toda esta infraestructura tecnológica no solo contribuye a facilitar las tareas y actividades como la documentación y obtención de información, sino que sirven como punta de lanza para la investigación sobre los cambios que se experimentan, tanto en el uso de herramientas en la educación como en los posibles métodos de enseñanza que se requieren para incorporar estos recursos.

Uno de los objetivos claves en la enseñanza de las redes de computadoras en nuestra carrera de grado, es la de transmitir fundamentos y conceptos básicos a los estudiantes de la Licenciatura en Análisis de Sistemas. Sin embargo, desde hace un tiempo venimos investigando como facilitar la relación entre la realidad y las teorías y modelos [8], [9], [10], es decir, entre lo concreto y lo abstracto.

Así, las computadoras personales, con la variedad de software que se ha desarrollado, tienen en sí mismas un gran potencial para mejorar el proceso de enseñanza-aprendizaje, a fines de que:

- El aprendizaje sea más interesante.
- El aprendizaje sea activo, no pasivo en las aulas.
- Los estudiantes estén más motivados.
- El aprendizaje sea al ritmo del estudiante en forma personalizada.
- La educación sea permanente.

El Departamento de Informática de la Facultad de Ciencias Exactas de la UNSa, ve a la simulación como una herramienta de apoyo pedagógico, que facilita la investigación y el desarrollo de una mejor calidad y pertinencia de la educación y de formación de saberes contextualizados entre teoría y práctica en el estudiante. Debido a esto, se está utilizando en diversas materias software de simulación.

Si bien existen numerosas herramientas y programas de simulación aplicables a la redes de computadoras y su utilización en el ámbito no académico data de un tiempo atrás, nuestra investigación nos permitirá analizar la influencia de simuladores en la enseñanza de la redes contextualizada en la universidad. Existen muchas investigaciones relacionadas con la influencia de software de simulación en otras áreas como la física, la química y la electrónica [1], [2], [3], que utilizaremos de referencia.

Líneas de Investigación, Desarrollo e Innovación

La línea de investigación, pretende evaluar la influencia de la utilización de software de simulación, en la enseñanza de contenidos de redes de computadoras, en el ámbito de las asignaturas relacionadas con la temática en la carrera de Licenciatura en Análisis de Sistemas de la UNSa y en el ámbito de cursos de postgrado organizados por el C.I.D.I.A. – Centro de Investigación y Desarrollo en Informática Aplicada que depende de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta. Para ello, se llevará a cabo un estudio aplicado, longitudinal y experimental que, asumiendo la complejidad del fenómeno educativo, considera aspectos cuantitativos y cualitativos.

Las preguntas que nos planteamos en este sentido son:

A. ¿Cómo influyen en la enseñanza y el aprendizaje, los trabajos realizados con un software de simulación? ¿Qué efectos causan los programas de simulación en lo que hace la motivación del alumno a aprender a resolver los problemas analizados en las clases teóricas?

B. ¿Cuánto pueden contribuir los trabajos sobre el simulador a que se reafirmen los conceptos teóricos y la comprensión de los fundamentos y funcionamiento en general de las redes?

C. ¿Los trabajos realizados con software de simulación le hace sentir la seguridad de quien ha explorado un tema acabadamente, especialmente con la posibilidad de variar arbitrariamente los valores de los componentes y variables?

D. Dado que el alumno puede diseñar e inventar topologías y experimentar con ellas ¿hasta qué punto favorece la creatividad?

E. ¿Hasta qué punto contribuyen los métodos de simulación a reemplazar un aprendizaje memorístico por aprendizaje significativo?

F. ¿En qué momento conviene aplicar estos métodos ¿Cuáles?, antes y/o después de los trabajos prácticos tradicionales? ¿Qué condiciones deben darse para la aplicación de estos métodos?

G. ¿Es deseable estructurar otras actividades prácticas utilizando métodos de simulación?

De lo expresado, se desprende que si bien muchas universidades ya han comenzado a aplicar este recurso, porque intuitivamente se reconoce su importancia, no se han encontrado muchas investigaciones didácticas realizadas en relación al mismo, y mucho menos asociadas a teorías cognitivas.

Dentro del paradigma cuantitativo se encuentran los diseños experimentales y los métodos estadísticos de contraste de hipótesis y de correlación entre variables, que serán aplicados en la investigación.

De lo expuesto en el marco teórico, del resultado de la búsqueda realizada y la formación y experiencia adquiridas, surgieron las siguientes hipótesis causales de investigación, en relación a las correlaciones aprendizaje-métodos de simulación:

- Hipótesis I: A igualdad de recursos utilizados y experiencias, el alumno aprende mejor, más significativamente, cuando complementa esas experiencias con métodos de simulación.
- Hipótesis II: El uso de software de simulación como complemento a los

prácticos tradicionales o de laboratorio, mejora el aprendizaje de conceptos y la comprensión.

Asimismo, la investigación persigue en el marco del paradigma cualitativo los siguientes objetivos:

- Realizar entrevistas y estudios de casos con los alumnos durante los períodos previstos.
- Diseñar actividades prácticas y de investigación adecuadas para ser realizadas con los simuladores.
- Fomentar el aprendizaje cooperativo entre los estudiantes.
- Introducir a los estudiantes de nivel universitario en la cultura científica actual, que concibe la simulación por computadora como una herramienta fundamental para el estudio, la investigación y la experimentación.
- Investigar sobre las posibilidades de integración de estas herramientas con plataformas de e-learning, como por ejemplo Moodle.
- Contrastar el resultado producto de la investigación con estudios similares en otras universidades [11], [12].

Resultados alcanzados

La primera etapa consistió en la preparación del material, guías de trabajos prácticos para la utilización de los simuladores y realización de pruebas piloto para validación del mismo.

En esta segunda etapa del proyecto se realizaron diversas experiencias con simulación en asignaturas de la Licenciatura en Análisis de Sistemas y cursos de postgrado de la Universidad Nacional de Salta y en la Universidad Católica de Salta a través del convenio firmado entre ambas instituciones.

A continuación se describen brevemente las experiencias realizadas a la fecha.

Año 2012. Experiencia con el simulador Packet Tracer en la asignatura Conectividad y Teleinformática de la Licenciatura en Análisis de Sistemas de la UNSa. El temario fue Direccionamiento IP. Protocolos de control. Ruteo estático y dinámico con RIP. El taller con simulación lo realizaron un grupo de alumnos voluntarios que cursaban la asignatura. Al finalizar, se solicitó a los alumnos que redacten su experiencia del uso del simulador. De la evaluación surgieron preguntas que sirvieron de base para la evaluación de las próximas experiencias.

Año 2012. Experiencia con el simulador KivaNS en la asignatura Conectividad y Teleinformática de la Licenciatura en Análisis de Sistemas de la UNSa. El temario fue concretamente el protocolo IP, ARP, subredes y ruteo estático. Lo realizaron un grupo voluntario de alumnos que cursaban la asignatura. Al finalizar el taller, se realizó una encuesta dividida en dos partes, en una primera parte se evaluó a KivaNS como herramienta de software y una segunda parte se evaluó KivaNS en el desarrollo de laboratorios y prácticas.

Año 2013. Experiencia con el simulador Packet Tracer en la asignatura Redes de Computadoras I de la Licenciatura en Análisis de Sistemas de la UNSa. El objetivo fue ambicioso, ya que se desarrollaron guías de trabajos prácticos para el simulador, para los temas introductorios de fundamentos y para cada protocolo tratado en la capa de aplicación, de manera de acompañar casi la mitad del cursado de la asignatura, con el uso del simulador. Lo realizaron un grupo de alumnos que cursaban la asignatura, seleccionados en forma aleatoria. Al finalizar cada tema, se realizó una encuesta. En primera instancia se evaluó a Packet Tracer como herramienta de software, luego se evaluó a Packet Trace y los Fundamentos de Redes, y finalmente Packet Tracer y la Capa de Aplicación.

Año 2013. Experiencia con el simulador Packet Tracer en un curso de Postgrado organizado por el CIDIA. Como temario se propuso una práctica a los estudiantes a fin de que pudieran realizar una configuración compleja jerárquica de servidores DNS (de múltiples niveles). El grupo de alumnos fue seleccionado en forma aleatoria. Al finalizar el taller, se realizó una encuesta dividida en dos partes, en una primera parte se evaluó a Packet Tracer como herramienta de software y una segunda parte se evaluó puntualmente a Packet Tracer y el protocolo DNS.

Año 2014. Experiencia con Packet Tracer en la asignatura Redes de Computadoras I de la Licenciatura en Análisis de Sistemas de la UNSa. Al temario ya incluido en esta asignatura, se sumaron guías prácticas para el simulador correspondientes a la capa de transporte y la capa de red, abarcando así todo el temario de la asignatura con el uso del simulador. Lo realizaron un grupo de alumnos que cursaban la asignatura, seleccionados en forma aleatoria. Al finalizar cada tema, se realizó una encuesta. En primera instancia se evaluó a Packet Tracer como herramienta de software, luego se evaluó a Packet Trace y los Fundamentos de Redes, Packet Tracer y la Capa de Aplicación, Packet Tracer y la Capa de Transporte y finalmente Packet Tracer y la Capa de Red. Además se realizaron encuestas para conocer la disponibilidad de equipos por parte de los alumnos y si trabajan o no, en esta instancia del cursado de la carrera.

Año 2014. Experiencia con el simulador MACSim en la asignatura Redes de Computadoras II de la Licenciatura en Análisis de Sistemas de la UNSa. La enseñanza se focalizó en redes LAN y puntualmente en el concepto del algoritmo de contención del medio CSMA/CD. Lo realizaron todos los alumnos que cursaban la asignatura pero a distintos tiempos divididos en dos grupos. Esta organización permitió evaluar el desempeño del rendimiento de los alumnos que realizaron el taller y contrastar objetivamente con el grupo que no realizó el taller en una primera instancia.

Año 2014. Experiencia con el simulador Packet Tracer y el Protocolo ARP en la asignatura Redes de Computadoras II de la Licenciatura en Análisis de Sistemas de la UNSa. La enseñanza se focalizó en el Protocolo ARP, resolución local en la misma LAN y resolución remota, cuando el destino está fuera de la LAN y se debe atravesar un router. Con el mismo esquema anterior, para que todos los alumnos puedan realizar el taller, pero que permita contrastar el desempeño del que estudia con simulación y el que estudia sin simulación.

Año 2014. Experiencia con el simulador MACSim en la asignatura Conectividad y Teleinformática de la Licenciatura en Análisis de Sistemas de la UNSa. La enseñanza se focalizó en redes LAN y puntualmente de concepto del algoritmo de contención del medio CSMA/CD. Lo realizaron todos los alumnos que cursaban la asignatura pero a distintos tiempos divididos en dos grupos. Esta organización permitió evaluar el desempeño del rendimiento de los alumnos que realizaron el taller y contrastar objetivamente con el grupo que no realizaron el taller en una primera instancia.

Año 2015. Experiencia con el simulador Packet Tracer y el Protocolo DNS en la asignatura Redes de Computadoras I de la Licenciatura en Análisis de Sistemas de la UNSa. La enseñanza se focalizó en el Protocolo DNS, resolución iterativa y recursiva. Se tomó un grupo de control para realizar la práctica con el simulador y contrastar con lo que realizaron la práctica tradicional en papel. Se evaluó el desempeño de los alumnos que realizaron el taller en contraste con los que no realizaron el taller con simulación.

Año 2015. Experiencia con GNS3 en un curso dictado en el marco de las Jornadas de Ingeniería de la UNSa. El temario fue ruteo estático y dinámico con el protocolo RIP. El taller con simulación lo realizaron todos los alumnos que se inscribieron al curso. Al finalizar el curso, se realizó una encuesta, en la cual se evaluó a GNS3 como herramienta de software.

Año 2015. Experiencia con el simulador Packet Tracer y el Direccionamiento IP en la asignatura Redes de Computadoras I de la Licenciatura en Análisis de Sistemas de la UNSa. La enseñanza se focalizó en el Direccionamiento IP, los conceptos de subredes IP, CIDR y máscara de red. Se contrastó objetivamente el desempeño de los alumnos que realizaron el taller con simulación y los que estudiaron de la forma tradicional con resolución de prácticos en papel.

Año 2015. Experiencia con el simulador Packet Tracer y el Ruteo IP en la asignatura Redes de Computadoras I de la Licenciatura en Análisis de Sistemas de la UNSa. La enseñanza se focalizó en el Ruteo IP, los conceptos de ruteo estático y ruteo dinámico con el protocolo RIP. Se contrastó objetivamente el desempeño de los alumnos que realizaron el taller con simulación y los que estudiaron de la forma tradicional con resolución de prácticos en papel.

Año 2015. Experiencia con el simulador MACSim en la asignatura Redes I de la Ingeniería en Telecomunicaciones en la Universidad Católica de Salta. Se replicó la experiencia realizada en el ámbito de la UNSa para la asignatura Redes de Computadoras II que tiene un programa similar. Se realizó un convenio de colaboración entre ambas universidades.

Año 2015. Experiencia con GNS3 en un curso de extensión organizado por el CIDIA. El temario fue ruteo estático y dinámico con el protocolo RIP. El taller con simulación lo realizaron todos los alumnos que se inscribieron al curso. Al finalizar el curso, se realizó una encuesta, en la cual se evaluó a GNS3 como herramienta de software.

Año 2015. Experiencia con el simulador MACSim en la asignatura Redes de Computadoras II de la Licenciatura en Análisis de Sistemas de la UNSa. La enseñanza se focalizó en redes LAN y puntualmente de concepto del algoritmo de contención del medio CSMA/CD. Lo realizaron todos los alumnos que cursaban la

asignatura pero a distintos tiempos, divididos en dos grupos. Esta organización permitió evaluar el desempeño del rendimiento de los alumnos que realizaron el taller y contrastar objetivamente con el grupo que no realizó el taller en una primera instancia.

Año 2015. Experiencia con el simulador Packet Tracer y el Protocolo ARP en la asignatura Redes de Computadoras II de la Licenciatura en Análisis de Sistemas de la UNSa. La enseñanza se focalizó en el Protocolo ARP, resolución local en la misma LAN y resolución remota cuando el destino está fuera de la LAN y se debe atravesar un router. Se mantuvo el mismo esquema anterior para que todos los alumnos puedan realizar el taller, pero que permita contrastar el desempeño del que estudia con simulación y el que estudia sin simulación.

Para el año 2016 está planificado replicar varias de las experiencias realizadas en el ámbito de la UNSa y UCASAL. Recientemente se firmó convenio de colaboración con la Universidad Nacional de San Juan, con el objeto diseñar en conjunto las experiencias con simulación a realizar en las asignaturas relacionadas con las redes de datos en las carreras Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas de Información

Además, se publicó un libro denominado “Redes de Computadoras I con Packet Tracer”, que recientemente fue evaluado y aprobado para su publicación en la Editorial de la Universidad Nacional de Salta - EDUNSa.

Se realizó un análisis estadístico descriptivo de cada experiencia realizada, y en términos generales se puede adelantar que fueron muy positivos. Sin embargo, no encontramos en este momento, realizando un análisis más exhaustivo que nos permita contrastar estadísticamente, si ambos grupos de estudio (con simulación y sin simulación) presentan una diferencia significativa en cuanto a alguna medida de tendencia central o de variabilidad, a los fines de poder tomar una decisión confiable sobre el beneficio o no

de la aplicación de la nueva técnica de enseñanza, es decir, si el uso de la herramienta de simulación como complemento a las prácticas tradicionales, favorece el aprendizaje. Lo ideal para este trabajo, sería comprobar estadísticamente que el método de enseñanza con simulación presenta ventajas significativas con respecto al método de enseñanza tradicional.

Formación de Recursos Humanos

El equipo de investigación se conforma de un director y cinco miembros entre docentes, egresados y estudiantes avanzados. Además se suman docentes de otras dependencias de la universidad con trabajos concretos de tesis de especialidad en seguridad y redes. También contamos con el apoyo de un asesor. Formación en postgrado

- Especialidad en Seguridad y Redes – UNLP. “Virtualización para la enseñanza de redes IP”. Expte. 3300-3489/11. Miguel A. Aguirre.
- Especialidad en Seguridad y Redes – UNLP. “Herramientas de Software de Simulación para Redes de Comunicaciones”. Expte. 3300-5305/12. Juan Antonio Torres. (Aprobada en julio de 2015).
- Doctorado en Ciencias Informáticas – UNLP. “Estudio de la influencia de un entorno de simulación en la enseñanza de redes en el nivel universitario”. Expte. 3300-5743/13. Daniel Arias Figueroa. En etapa de elaboración de tesis con todos los créditos de cursos convalidados.
- Magister en Redes de Datos – UNLP. “Un estudio comparativo en extensiones de seguridad para el sistema de nombre de dominio (DNS)”. Expte. 3300-002113/10-000. Ernesto Sánchez.
- Magister en Ingeniería de Software – UNLP. “Desarrollo basado en Conocimiento siguiendo prácticas ágiles”. Expte. 3300-006507/13-000. Loraine Gimson. (Aprobada en noviembre de 2015).

- Cursando la Maestría en Redes de Datos en la UNLP. Ing. Alvaro Gamarra.
- Cursando la Licenciatura en Análisis de Sistemas en la UNSa. Diego Gil.

Referencias

[1]. Trabajos Prácticos, métodos de simulación y Aprendizaje significativo. Edgardo Cámara – Gloria E. Alzugaray. Universidad Tecnológica Nacional. Facultad Regional Santa Fe. Aplicado a la asignatura Electrónica II.

[2]. “Tecnologías de Información y Comunicación como medios de aprendizaje de Redes de Computadoras”. <http://es.scribd.com/doc/17335747/Tesis-Utilizacion-de-Tics-para-el-aprendizaje-de-Redes-de-Computadoras>

[3]. Estudio de la influencia de un entorno de simulación por ordenador en el aprendizaje por investigación de la física en bachillerato. <http://books.google.com.ar/books?id=IWCQmq-20E.S.O.%20y%20bachillerato&f=false>

[4]. Objetos educativos abiertos, la simulación en software libre. <http://www.cibersociedad.net/congres2009/es/coms/objetos-educativos-abiertos-la-simulacion-en-software-libre/341/>

[5]. Castillo C. y Arrieta X. Referentes teóricos para el diseño y evaluación de software de apoyo a la enseñanza – aprendizaje de la física. Memorias de la IX Conferencia Interamericana de Educación en Física CIAEF-2006- IACPE. San José de Costa Rica. 3 a 7 de julio de 2006. <http://www.efis.ucr.ac.cr/varios/ponencias/9referentes%20teoricos.pdf>

[6]. Díaz-Antón, G. (2002) *Uso de software educativo de calidad como herramientas de apoyo para el aprendizaje*. Jornadas educativas: “La escuela como instrumento de cambio”, IEA, Abril, Caracas. <http://www.academia-interactiva.com/articulos.html>

[7]. Franco I, Álvarez F. *Los Simuladores, estrategia formativa en ambientes virtuales de aprendizaje*. Revista Virtual Universidad Católica ISSN 0124-5821. http://www.ucn.edu.co/portal/uzine/volumen21/articulos/3_Investigaci%C3%B3n_simuladores.pdf

[8]. "KIVA: Un simulador para la enseñanza de Redes IP" VII Congreso Tecnología en Educación y Educación en Tecnología (TE&ET 2012) Ciudad de Pergamino, Provincia de Buenos Aires – Junio de 2012. Daniel Arias Figueroa, Gustavo D. Gil, Ernesto Sánchez (C.I.D.I.A).

[9]. "Una experiencia con simuladores en la asignatura Conectividad y Teleinformática" I Jornadas Nacionales de TIC e Innovación en el aula – III Jornadas de Experiencias en EaD de la UNLP - Universidad Nacional de La Plata. La Plata. Diciembre de 2011. Daniel Arias Figueroa, Gustavo Gil, Sergio Rocabado Moreno, Ernesto Sánchez, Gabriela Salvadó (C.I.D.I.A)

[10]. "Una experiencia en el uso de software de captura de tráfico y virtualización como apoyo a la enseñanza de redes de datos" I Jornadas Nacionales de TIC e Innovación en el aula – III Jornadas de Experiencias en EaD de la UNLP - Universidad Nacional de La Plata. La Plata. Diciembre de 2011. Ernesto Sánchez, Daniel Arias Figueroa, Sergio Rocabado Moreno (C.I.D.I.A).

[11]. Teaching Computer Networks through Network Simulation Programs – University of Derby - Faculty of Business, Computing and Law – School of Computing. http://www.derby.ac.uk/files/shao_ying_zhu_-_teaching_computer_networks_through_network_simulation_programs.pdf

[12]. Improving a Computer Networks Course Using the Partov Simulation Engine Behnam Momeni and Mehdi Kharrazi. <http://sharif.edu/~kharrazi/pubs/te12.pdf>

Técnicas provenientes de las ciencias de la computación aplicadas a la simulación macroeconómica

Francisco Ibañez, Daniel Díaz, Sandra Oviedo, Nancy Alonso, Fabián Saffe
LISI- Instituto de Informática – Dpto. de Informática
FCEfyN - Universidad Nacional de San Juan
CUIM – Av. Ignacio de la Roza 590 (O), Rivadavia – J5402DCS San Juan
{ fibanez, ddiaz, soviedo, nalonso, fsaffe}@iinfo.unsj.edu.ar}

Resumen

Este trabajo describe los progresos de la construcción de un simulador de escenarios macroeconómicos, con la complejidad matemática de los modelos, reflejados en la implementación del software, pero con una interface gráfica, que facilite al usuario del simulador la interpretación de los resultados, con la finalidad de aumentar la cantidad de personas que puedan familiarizarse con la naturaleza de los problemas económicos.

Palabras clave:

Simulación, Macroeconomía, Educación

Contexto

Uno de los aportes de la simulación, es contribuir a la comprensión del comportamiento de un sistema. El propósito de este trabajo es describir los progresos que se están llevando a cabo en la investigación encuadrada en el proyecto “Propuesta interdisciplinaria para la construcción de un simulador macroeconómico”, que se desarrolla en el Laboratorio de Informática Aplicada a la Innovación del Instituto de informática de la Facultad de Ciencias Exactas Físicas y

Naturales de la Universidad Nacional de San Juan.

Introducción

La simulación es uno de los enfoques utilizados para abordar los problemas macroeconómicos. Según Robert E. Shannon [1]. “La Simulación es el proceso de diseñar y desarrollar un modelo computarizado de un sistema o proceso y conducir experimentos con este modelo, a fin de entender el comportamiento del sistema o evaluar varias estrategias con las cuales se puede operar el sistema”. La principal ventaja de la utilización de simuladores es la reducción de costos, tiempo y riesgos en comparación con la experimentación real. Según [2] una de las funciones de la teoría económica es proveer sistemas económicos artificiales totalmente articulados que sirvan como laboratorios para experimentar políticas que serian prohibitivamente costosas de experimentar en economías reales.

Por otro lado, la principal desventaja de los simuladores en general, es que pueden ocasionar resultados que estén lejos de la realidad, debido a las diferentes interpretaciones de la información y a las limitaciones de los modelos. Sin embargo, a partir de la definición de Shanon, podemos resaltar que en el caso de que la simulación no se corresponda con la realidad en una

medida aceptable, aun puede ayudar a “entender el comportamiento del sistema”.

En el campo de la economía, incluso cuando los modelos utilizados en los simuladores no constituyen una gran teoría unificada, los modelos pueden constituir un conjunto de herramientas que pueden mejorar el mundo, tal cual se expresa en [3].

La simulación, como herramienta para comprender las relaciones de causalidad subyacentes a los modelos económicos, constituye el tema central que nos interesa abordar en este proyecto. Especialmente para el caso particular de la Argentina.

Los conocimientos de Macroeconomía, y la contribución de un simulador, no sólo son importantes para decidir medidas económicas, por parte de las autoridades, sino para que en tiempos de elecciones, los partidos políticos expliquen las medidas económicas que proponen y el pueblo pueda elegir el voto, basado en una comprensión más profunda de las propuestas, de los impactos económicos, sociales y medioambientales, no solamente en Argentina, sino en el mundo entero. Tomar una decisión sin conocimiento, puede conducir a un estado, en donde el pueblo termina preso de su propia elección. Reivindicando la frase histórica de Sócrates, “El conocimiento os hará libres”.

El desafío radica en la dificultad de divulgar un conocimiento científico basado no sólo en conceptos provenientes de áreas humanísticas, sino también en modelos matemáticos, que pierden casi toda su esencia, cuando el único recurso que se usa para explicarlos son las palabras.

Líneas de Investigación, Desarrollo e Innovación

Este trabajo de investigación presenta cuatro ejes, el macroeconómico, el matemático, el de las ciencias informáticas, y el pedagógico. Desde el punto de la macroeconomía, la principal contribución que se propone consiste en la implementación y uso de la técnica “paso a paso” proveniente del contexto del aprendizaje y enseñanza de compiladores. Esta técnica permite visualizar paso a paso la secuencia de gráficos que conduce al resultado. Actualmente se dispone de una herramienta para el aprendizaje y enseñanza de compiladores en la cual se ha implementado tal técnica [4] [5], y también se ha desarrollado un prototipo inicial del simulador macroeconómico, basado en la técnica “paso a paso”.

Desde el punto de la Informática, en el desarrollo futuro del simulador, proponemos el uso de técnicas provenientes de la Inteligencia Artificial, tales como algoritmos genéticos, para resolver problemas de maximización de funciones con extremos ligados, que no pueden resolverse con métodos provenientes del análisis matemático clásico, como por ejemplo, el método de Lagrange. A los efectos de implementar mediante simulación, el modelo Nuevo Keynesiano, es necesario resolver el problema de maximización de funciones con extremos ligados [6]. Los algoritmos genéticos son adecuados, en los casos en que las condiciones necesarias para aplicar Lagrange no se cumplen [7].

Desde el punto vista pedagógico, este proyecto pretende fortalecer la enseñanza de los fenómenos macroeconómicos mediante la incorporación y uso del simulador en el proceso

enseñanza/aprendizaje de Macroeconomía.

En cuanto al equipo de trabajo, la construcción de un simulador macroeconómico tal cual se ha planteado necesita ser multidisciplinario. Es por ello

que este proyecto pretende integrar docentes e investigadores de las ciencias sociales, exactas e informática. Esta integración contribuirá a fortalecer el conocimiento de cada área en particular.

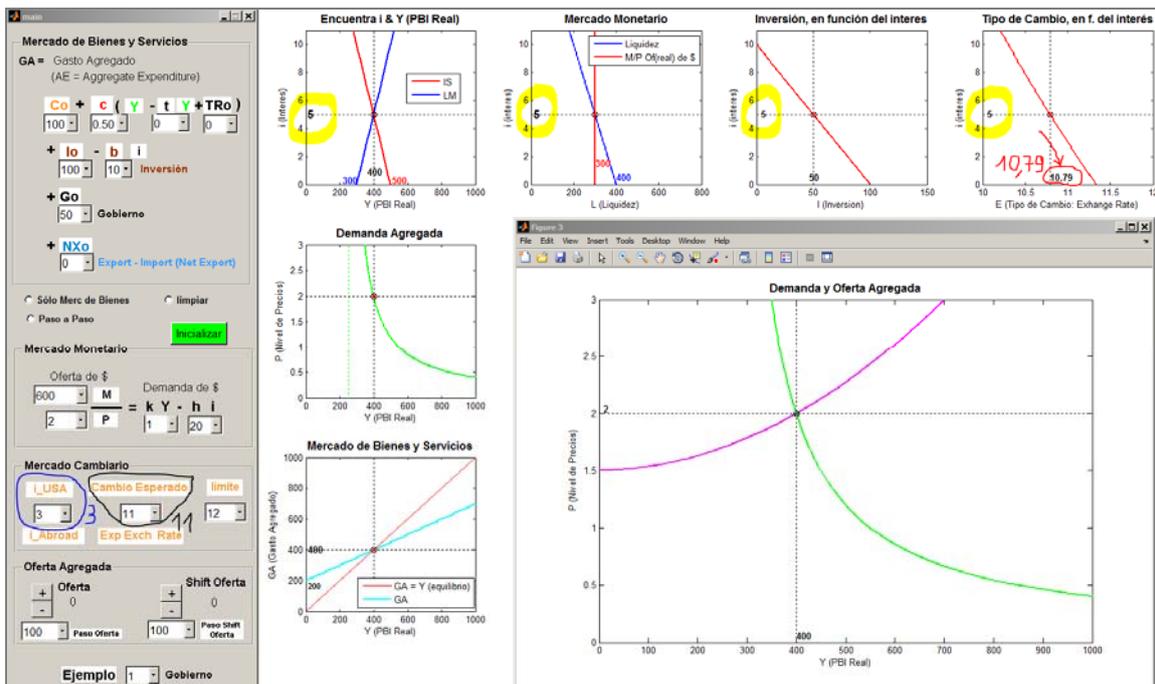


Figura 1: Simulación de la determinación del tipo de cambio

Resultados y Objetivos

Aunque el trabajo se encuentra en su etapa inicial, existe en funcionamiento un prototipo básico con el cual se ha desarrollado un material pedagógico.

En [8] se pueden ver los resultados de una primera versión del simulador. Aquí sólo incluiremos una explicación sucinta de esta versión.

En el ejemplo de la figura 1, partimos de determinados valores asignados a todas las variables exógenas. Las variables que nos interesan en este apartado son dos, la tasa de interés en Estados Unidos, marcada en el gráfico con azul, que tiene asignado el valor de 3, (3 % anual), y el tipo de cambio que se

espera en Argentina en un año a partir de la fecha actual, marcada con negro, que tiene asignado el valor 11 (11\$ por dólar). Además de estas dos variables exógenas, la otra variable que determina el equilibrio en el mercado cambiario, que es el tipo de interés en Argentina, es endógena, es decir, queda determinada por el simulador a partir de los valores de variables exógenas. Para el caso del ejemplo, el simulador determina que, con todos los mercados en equilibrio, el tipo de interés es de 5 (5% anual), marcado con color amarillo, y el tipo de cambio es de 10.79\$ por dólar, marcado con color rojo en la esquina superior derecha.

En el "1º Encuentro Internacional de Investigadores de Ciencias Económicas", desarrollado en San Juan, se presentó una

segunda versión del simulador, la cual puede verse en [9].

Formación de Recursos Humanos

El equipo de trabajo está constituido por tres docentes del área de informática, uno del área de economía, uno del área de matemáticas, y dos alumnos de grado.

En esta etapa, el equipo está abocado a la tarea de difundir la temática entre los alumnos de grado y posgrado que realicen sus tesis en el contexto de esta línea de investigación.

Referencias

- [1] R. E. Shannon, "Introduction to the art and science of simulation," in *Proceedings of the 30th conference on Winter simulation*, 1998, pp. 7-14.
- [2] R. E. Lucas, "Methods and problems in business cycle theory," *Journal of Money, Credit and banking*, pp. 696-715, 1980.
- [3] L. Maxted, "Economics Rules: The Rights and Wrongs of the Dismal Science," ed: REED BUSINESS INFORMATION 360 PARK AVENUE SOUTH, NEW YORK, NY 10010 USA, 2015.
- [4] F. Ibañez, D. Diaz, S. Oviedo, and A. Otazú, "COMPI, una herramienta interactiva para la enseñanza de construcción de compiladores," presented at the Workshop de Investigadores en Ciencias de la Computación (WICC 2015), Salta - Argentina, 2015.
- [5] F. Ibañez, D. Diaz, S. Oviedo, and A. Otazú, "Herramienta Educativa para la Construcción de Compiladores," presented at the Congreso Internacional de Educación, Tecnología y Ciencia (CIETyC 2015) La Guajira - Colombia 2015.
- [6] Y. Ahmad, "The Basic Structure of New Keynesian Models. Deriving the New Keynesian IS Curve," *Macroeconomics course university wisconsin whitewater*, vol. http://facstaff.uww.edu/ahmady/courses/econ402/lectures/nk_is_lecture.pdf, 2015.
- [7] J. E. M. Legaz, "La interpretación económica de los multiplicadores de Lagrange," *Matematicalia: revista digital de divulgación matemática de la Real Sociedad Matemática Española*, vol. 2, p. 4, 2006.
- [8] F. Ibañez, "Propuesta de un proyecto interdisciplinario para construir un simulador macroeconómico," in *JATIC 2015*, Mar del Plata, 2015.
- [9] F. Ibañez, "<https://www.youtube.com/watch?v=BTuPpxpyMPM>," *1er encuentro internacional de investigadores de ciencias económicas*, vol. UNSJ, 2015.

Evaluación de aprendizajes: instrumentos e interactividad en entornos virtuales

Vilanova Gabriela -Lezcano, Laura Noemí

Universidad de la Patagonia Austral /Proyecto de Investigación 29/B177,
Aprender y enseñar con las tics como instrumentos mediadores en los procesos
de construcción de conocimiento del Instituto de Educación y Ciudadanía (IEC)
Unidad Caleta Olivia-Santa Cruz
vilanova@uolsinectis.com.ar y lezcanolaura@gmail.com

Resumen

La evaluación es un elemento constitutivo de los procesos de enseñanza y de aprendizaje, y como tal, resulta fundamental para producir información que les permita a los estudiantes reconocer aciertos y dificultades en su proceso y actuar en consecuencia, implementando estrategias de estudio y aprendizaje. Al mismo tiempo, brinda información vital para que el docente reoriente su diseño de enseñanza. Desde esta perspectiva, la evaluación se realiza en diferentes momentos del proceso educativo y sobre diferentes elementos y situaciones.

La influencia de la tecnología en los modos de aprender genera nuevos desafíos para la evaluación de aprendizajes.

A partir del diseño, la implementación y evaluación de propuestas de formación en los entornos virtuales por parte del profesor y en co-gestión con el estudiante, se pueden potenciar buenas prácticas de aprendizaje y de construcción conjunta de conocimiento.

Palabras clave: evaluación de aprendizajes, entornos virtuales, instrumentos, interactividad.

Contexto

En el marco del Proyecto de Investigación 29/B177, Aprender y enseñar con las tecnologías de la información y la comunicación como instrumentos mediadores en los procesos de construcción de conocimiento del Instituto de Educación y Ciudadanía (IEC) de la Universidad de la Patagonia Austral (UNPA), se inicia en el 2015 una línea de investigación sobre evaluación de aprendizajes en entornos virtuales.

Como estudiante de pos grado en proceso de elaboración de tesis para la Maestría en Educación en Entornos Virtuales de la UNPA, la participación en el proyecto gira en torno al análisis de las semejanzas y diferencias entre la evaluación de aprendizaje en un entorno virtual y en un entorno presencial, como así también los instrumentos de evaluación de aprendizajes en entornos virtuales como facilitadores de la interactividad.

Introducción

1.1 ¿Semejanzas y/o diferencias entre la evaluación de aprendizajes en entornos virtuales y los entornos presenciales?

La evaluación se considera como:

- un proceso constitutivo de cualquier propuesta didáctica;
- una oportunidad para promover una permanente reflexión sobre logros y obstáculos en el desarrollo de la tarea,
- un proceso sistemático, continuo e integrado dentro de un proceso más general que es el educativo.

“...Defino la evaluación como un proceso que, a partir del conocimiento y comprensión de cierta información, permite emitir un juicio de valor acerca de un aspecto de la realidad en la cual se interviene en un determinado contexto socio histórico particular y que, a la vez que posibilita tomar decisiones, exige desde el diálogo con quien esté involucrado, argumentar justificaciones del juicio de valor realizado...” [1]

La cuestión de la evaluación relacionada al uso de tecnología y la comunicación no se refiere sólo a transferir modelos de evaluación tradicionales a entornos virtuales o analizar si es necesario generar nuevas concepciones relacionadas con los propósitos y formas de evaluación. Lo que debe interesar son los fundamentos pedagógicos que sustentan una evaluación auténtica en los entornos virtuales, qué oportunidades de aprendizaje y evaluación se pueden crear mediante tecnología de la información y la comunicación, o en qué aspectos se enriquecen o empobrecen actuaciones tanto de los profesores como de los

estudiantes. El proceso de evaluar atraviesa los procesos de enseñanza y de aprendizaje de manera tal, que si se analizara y modificara en profundidad la idea de evaluar se modificarían sustancialmente los procesos de intervención de los docentes. [2]

Para identificar el avance de los procesos cognitivos e interactivos en entornos virtuales de aprendizaje, es necesario contar con una estrategia flexible de evaluación así como una serie de herramientas que permitan detectar cambios en la complejidad de las construcciones de conocimientos por parte de los estudiantes. El rol que la evaluación tiene en la formación on line obliga a una planificación y ejecución minuciosa y el análisis atento de cada uno de sus elementos: cada evaluación conlleva una retroalimentación casi inmediata.

La etapa de búsqueda, conocimiento y comprensión de la información tiene un papel fundamental en el proceso de evaluación. El correcto empleo de los instrumentos adecuados permitirá tener información válida sobre el objeto seleccionado. Los juicios y decisiones que posteriormente se tomen dependen de la información brindada por ellos. Los instrumentos que se emplean para la etapa de relevamiento de información son parte del proceso de enseñanza y es allí donde cobran sentido. Según la información que deseamos obtener y manteniendo la coherencia de la estructura con la finalidad que se persigue se seleccionarán estos instrumentos. En este sentido los instrumentos de evaluación de aprendizajes poseen un valor agregado como elementos facilitadores de la interacción entre docentes y estudiante.

1.2 Instrumentos de evaluación e interactividad

La evaluación y la interactividad se relacionan, desde el momento en que las evaluaciones son interactivas, implican retroalimentación a y con los estudiantes, en ocasiones es andamiaje de sus avances y progresos. Las evaluaciones formativas son ejemplos de trabajo interactivo.

El concepto de interactividad es complejo y polisémico, sin embargo se lo puede comprender como “... *es un diálogo, discurso o evento entre dos o más participantes y objetos que ocurre sincrónica o asincrónicamente, mediado por respuesta o retroalimentación, teniendo a la tecnología como interfaz...*”. [3] Desde esta perspectiva se distingue el concepto de interacción del concepto de interactividad en la educación on line. Esta última comprende la forma, la función y el impacto de las interacciones en la enseñanza y el aprendizaje. Se concibe como un proceso, como reciprocidad o diálogo. Dentro del marco de la interactividad, se pueden contemplar sus productos como significados construidos socialmente. Se generan, así construcciones nuevas a partir de una tarea y de la interacción con otros estudiantes y/o con los docentes

La consideración de la interactividad es central para poder comprender los procesos de construcción de conocimientos. Es el núcleo en torno al cual se construyen los conocimiento nuevos, consta de una serie de intercambios entre los participantes que permiten, en el contexto de un entorno virtual de formación, procesar la formación relevante, obtener retroalimentación de otros participantes (estudiantes- pares-docentes) y, o bien,

ajustar la construcción o bien generando nuevos conocimientos emergentes.

En este contexto la estrategia e instrumentos de evaluación de aprendizaje asumen particular relevancia como facilitadores de la interacción entre quienes participan. Ejemplo de ello son los foros, los e-portfolios y las rúbricas, entre otros.

Dadas las características de la evaluación de aprendizajes en entornos virtuales, donde pueden desarrollarse algunas limitaciones en la comunicación impuestas por la naturaleza del medio, la evaluación continua del aprendizaje es esencial, su función de retroalimentación al estudiante sobre sus progresos avances u obstáculos.

Líneas de Investigación, Desarrollo e Innovación

La participación en el proyecto comenzó con preguntas acerca de la evaluación de aprendizajes en entornos virtuales en espacios de formación y sobre cómo los instrumentos de evaluación pueden convertirse en facilitadores de la interactividad entre estudiantes y docentes en beneficio de aquellos.

Cada interrogante dio lugar a otras preguntas: ¿el proceso de evaluación de aprendizajes en ambientes presenciales es aplicable a ambientes virtuales?, ¿se modifican las particularidades de la evaluación de aprendizajes al cambiar estrategias y entornos de aprendizaje?, ¿cuál es en este camino la relación entre evaluación, instrumentos, interactividad y construcción de conocimientos? ¿Cuál es la perspectiva de los estudiantes? ¿Cómo

viven el proceso de formación en entornos virtuales? ¿Cómo valoran la evaluación de sus propios aprendizajes?

En forma gradual se avanzó en la revisión de los antecedentes y en el relevamiento de la perspectiva de estudiantes. Son estos avances los que se desarrollan en esta presentación y que se organizan de la siguiente manera:

- Principios y antecedentes que encuadran el posicionamiento sobre la evaluación de aprendizajes en entornos virtuales.
- Avances en el relevamiento de las perspectivas de estudiantes sobre la modalidad virtual y la propuesta de evaluación de sus aprendizajes.

En este proceso que no se ha cerrado, ni finalizado, se avanzará en la búsqueda de nuevas respuestas que nos permitan interpelar y mejorar las intervenciones docentes cotidianas en los ámbitos de formación en entornos virtuales.

Resultados y Objetivos

En el diseño de propuestas mediadas por tecnología o desarrolladas en entornos virtuales, se considera que el docente posee un rol mediador, ya que facilita la construcción del conocimiento de los estudiantes y su aplicación. Sin la intervención de los docentes no es posible hablar de aportes, de obstaculizadores o facilitadores de la tecnología.

Ahora bien: ¿qué perspectiva poseen los estudiantes sobre la evaluación de sus propios aprendizajes en estos entornos?

En este recorrido se han logrado algunas precisiones como las mencionadas hasta el momento y se estima oportuno continuar con la exploración y sondeo de la perspectiva de los estudiantes que han tenido alguna experiencia como tales en entornos virtuales.

Los próximos pasos serán:

- Realizar un sondeo de las valoraciones de los estudiantes sobre el entorno para rescatar aspectos que puedan aportar elementos a la mejora de cualquier propuesta y en particular lo referido a la evaluación de aprendizajes.
- Avanzar en el análisis y reflexión sobre las prácticas cotidianas de evaluación de aprendizajes.

El entorno favorece las posibilidades de interactividad ofreciendo una serie de recursos y herramientas en las plataformas, entonces ¿cuáles son los criterios de los docentes en sus elecciones? ¿Existen aspectos que obturen la intencionalidad pedagógica y didáctica del profesor para generar propuestas de evaluación que propicien la interactividad y que favorezcan el aprendizaje de los estudiantes? ¿Cuáles son esas variables?: ¿escasa formación? ¿Prácticas de la presencialidad arraigadas en las intervenciones en el entorno virtual?

Formación de Recursos Humanos

El equipo del Proyecto está constituido por perfiles académicos diversos que enriquecen las líneas de investigación propuestas: Licenciada en Ciencias de la Computación, Ingeniero en Seguridad e Higiene, Licenciada en Educación, Licenciados en Matemática,

cursantes de Maestría en Ingeniería en Sistemas y tesis de la Maestría en Educación en Entornos Virtuales.

En síntesis el equipo está conformado por docentes, investigadores adjuntos y auxiliares del Departamento de Ciencias Exactas y Naturales de Unidad Académica Caleta Olivia del Área Sistemas, Matemática y Seguridad e Higiene.

Referencias

- [1] Steiman J. (2008). Más didáctica en la educación superior. Miño y Dávila.p. 142-143
- [2] Camillioni, A. y otros. (1998) La evaluación de los aprendizajes en el debate didáctico contemporáneo. Bs. As.: Paidós.
- [3] Muirhead, B.; Juwah, C. (2004). Interactivity in computer-mediated college and university education: A recent review of the literature. *Educational Technology & Society*, 7. P. 13

Otras:

-Barberá, E. (2006). Aportaciones de la tecnología a la e-Evaluación. RED. Revista de Educación a Distancia, número especial VI. Consultado el 27/08/15 en <http://www.um.es/ead/red/M6>

-Arango, M. (2004). Foros Virtuales como estrategia de aprendizaje. Revista Debates Latinoamericanos N° 2.

- Garcia Aretio L (2001) La educación a distancia: de la teoría a la práctica. Barcelona: Ariel.

-Harasim, L.; Hiltz, S. R.; Turoff, M; Teles, L. (2000). Redes de aprendizaje. Guía para la enseñanza y el aprendizaje en red. Barcelona: Gedisa.

-Litwin E. (2005). Tecnología educativa en tiempos de internet. Bs.As. Amorrortu

-Romero, M. (2008). Implementación de la competencia transversal «Uso y aplicación de las TIC en el ámbito académico y profesional» en el contexto universitario de la UOC. *Relatec-Revista Latinoamericana De Tecnología Educativa*, 7(2), 81-89.

-Salinas, J., Perez, A., De Benito, B. (2008) Metodologías centradas en el alumno para el aprendizaje en red. Editorial Síntesis.

Hacia un modelo predictivo de rendimiento académico utilizando minería de datos en la UTN – FRRe

David L. la Red Martínez, Marcelo Karanik, Mirtha Giovannini, Reinaldo Scappini

Grupo de Investigación Educativa / Departamento de Ingeniería en Sistemas de Información / Facultad Regional Resistencia / Universidad Tecnológica Nacional
French 414, (3500) Resistencia, Argentina, +54-379-4638194

laredmartinez@gigared.com mkaranik@gmail.com meg_c51@yahoo.com.ar rscappini@gmail.com

Resumen

Durante el cursado de la carrera de Ingeniería en Sistemas de Información en la Facultad Regional Resistencia de la Universidad Tecnológica Nacional (UTN-FRRe), el alumno se enfrenta con la necesidad de cursar y regularizar ciertas materias que le generan restricciones de correlatividad. Ese es el caso de la cátedra de primer año Algoritmos y Estructuras de Datos, cuya regularización es necesaria para cursar varias asignaturas de segundo y tercer año. Tomando como base los resultados del proyecto “Determinación de perfiles de estudiantes y de rendimiento académico mediante la utilización de minería de datos”, 25/L059 - UTI1719, implementado en la mencionada cátedra (2013-2015), se está comenzando un nuevo proyecto que tiene como objetivo tomar como punto de partida el análisis descriptivo (lo que pasó), y utilizar la analítica avanzada, con el objetivo de explicar el por qué, el qué va a pasar y cómo podemos abordarlo. Para el estudio se utilizarán distintas herramientas de Minería de Datos: clustering, redes neuronales, redes bayesianas, árboles de decisión, regresión y series temporales, etc. Estas herramientas permiten obtener resultados desde distintas perspectivas del problema abordado. De esta manera se podrán detectar situaciones problemáticas potenciales al inicio del cursado y tomar las medidas necesarias para solucionarlas.

Palabras clave: rendimiento académico; almacenes de datos; minería de datos; modelos predictivos.

Contexto

En búsqueda de obtener respuestas al alto grado de desgranamiento en los pri-

meros años, se propuso el desarrollo del proyecto “Determinación de perfiles de estudiantes y de rendimiento académico mediante la utilización de minería de datos”, 25/L059 - UTI1719, implementado en la cátedra “Algoritmos y Estructuras de Datos” de la Ingeniería en Sistemas de Información de la UTN – Regional Resistencia (2013-2015). Como consecuencia de ello se establecieron los perfiles de rendimiento académico de los alumnos según su condición social, familiar, antecedentes académicos, etc., relacionándolos con su rendimiento académico en la asignatura mencionada.

Como continuación del proyecto mencionado se ha iniciado el proyecto “Diseño de un modelo predictivo de rendimiento académico mediante la utilización de minería de datos”, UTI3808TC (2016-2018), inserto en el ámbito del Grupo de Investigación Educativa (GIE) de la FRRe. El proyecto es financiado por la UTN, homologado e inserto en el programa de incentivos del Ministerio de Educación.

Introducción

La Universidad enfrenta actualmente el desafío de mejorar su calidad académica enfocándose no sólo en el sistema de enseñanza-aprendizaje, sino contemplando otras variables, como la sistematiza-

ción de procesos de evaluación permanentes (Briand et. al., 1999). Entre estas variables, se destaca el estudio del perfil de rendimiento académico de los estudiantes.

Se define al rendimiento académico como la productividad del sujeto, matizado por sus actividades, rasgos y la percepción más o menos correcta de los cometidos asignados (Maletic et. al., 2002).

Generalmente al evaluar el rendimiento académico, se analizan en mayor o menor medida los elementos que influyen en el desempeño como ser, entre otros, factores socioeconómicos, la amplitud de programas de estudio, las metodologías de enseñanza, conocimientos previos del alumno (Marcus, 2003).

Se ha demostrado con varios estudios que el factor más relacionado con la calidad educativa es el propio alumno como co-productor, medido a través del nivel socioeconómico del hogar de donde proviene (Maradona y Calderón, 2007) y se ha evidenciado que la productividad del estudiante es mayor para las mujeres, para los estudiantes de menor edad y para quienes provienen de hogares con padres más educados (Porto & Di Gresia, 2003).

También se ha mostrado el contraste que hay entre las personas que trabajan y estudian y las que solamente estudian, encontrándose que no existen diferencias en el rendimiento académico de los dos conjuntos (Reyes, 2004).

El problema de encontrar buenos predictores del rendimiento futuro de manera que se reduzca el fracaso académico en los programas de postgrado ha recibido una especial atención en EE. UU. (Wilson & Hardgrave, 1995), habiéndose encontrado que las técnicas de clasificación como el análisis discriminante o la regresión logística son más adecuadas que la regresión lineal múltiple a la hora de predecir el éxito/fracaso académico.

La diversidad de estudios sobre el rendimiento académico muestra que no existe una manera única para evaluarlo. Por ello, la determinación de grupos o clases de alumnos es un elemento a tener en cuenta para establecer las causas de los problemas relacionados al desempeño de los mismos. Más aún, los problemas pueden variar dependiendo el contexto regional y la realidad social donde está inserto el alumno. Es decir, no existen herramientas que se puedan aplicar a todos los ámbitos y los resultados tampoco pueden ser extensibles para explicar todas las situaciones posibles. Esto denota claramente la necesidad de determinar perfiles en las instituciones educativas específicas adaptando las herramientas a cada situación particular.

Surge, entonces, la necesidad de implementar un mecanismo que permita determinar las características propias del estudiante, analizando la existencia de relaciones y patrones de comportamiento estudiantiles que posibilite la definición clara de perfiles de alumnos. Para ello una alternativa es utilizar técnicas de minería de datos para el modelado descriptivo (La Red Martínez et. al., 2014a, 2014b, 2015).

A su vez, el modelado predictivo puede usarse para analizar una base de datos y determinar ciertas características esenciales acerca del conjunto de datos que permitan predecir el comportamiento de alguna variable (Connolly & Begg, 2005).

Líneas de Investigación, Desarrollo e Innovación

En la UTN-FRRe se dicta la carrera de Ingeniería en Sistemas de Información (ISI). Esta carrera tiene varias asignaturas en el primer año que son específicas a la profesión del Ingeniero en Sistemas, las cuales por lo general son las que generan

el mayor desgranamiento. Una de estas asignaturas es Algoritmos y Estructura de Datos.

En el marco del proyecto “Determinación de perfiles de estudiantes y de rendimiento académico mediante la utilización de minería de datos”, se trabajó en la identificación de las variables que explican el desigual rendimiento académico por parte de los estudiantes de la citada asignatura, lográndose modelos descriptivos del rendimiento académico; en el contexto del proyecto “Diseño de un modelo predictivo de rendimiento académico mediante la utilización de minería de datos” iniciado este año se buscará desarrollar modelos predictivos de rendimiento académico.

Teniendo en cuenta los resultados logrados en las evaluaciones realizadas durante el cursado de la asignatura, se buscó determinar en qué medida el desigual rendimiento académico es influenciado por variables socioeconómicas y actitudinales tales como: escuela media de procedencia, nivel educativo de los padres, nivel socioeconómico, edad, género, actitud general hacia el estudio, uso de herramientas de apoyo (campus virtual) (La Red Martínez et. al., 2015).

El universo estuvo constituido por los alumnos en condiciones de cursar la asignatura durante el año 2013, 2014 y 2015 y la unidad de análisis fue cada uno de esos alumnos.

El análisis de los resultados se basó en considerar como parámetro de minería la situación final del alumno, la cual refleja su estado en la materia al cierre del Ciclo Lectivo. Se consideró en situación de Libre al alumno que no aprobó ni los exámenes parciales ni los recuperatorios; Regular a quien logró aprobar los 3 exámenes parciales (recuperándolos o no) con nota mayor o igual al 60% pero que no alcanzó al menos el 75% en todos ellos y Promocionado a aquel que aprobó

todos los parciales con nota mayor o igual al 75%.

Resultados y Objetivos

Resultados

Se mostrarán a continuación algunos de los resultados del proyecto anterior que constituyen el punto de partida del proyecto que se inicia.

Se han obtenido los siguientes resultados: 81.42% de alumnos en condición de Libre, 10.62% Regular y sólo 7.96% Promocionado.

En los comentarios siguientes se considerará rendimiento académico “alto” al logrado por los alumnos con situación final de “promocionado”, rendimiento “medio” al de los alumnos con situación de “regular” y rendimiento “bajo” al de los alumnos con situación de “libre”; a su vez se considerará “éxito académico” al rendimiento “alto” y “medio” y “fracaso académico” al rendimiento “bajo”.

Seguidamente se comentarán algunos de los aspectos que se considera oportuno destacar:

- Teniendo en cuenta el tipo de escuela secundaria de la cual provienen los alumnos, se observó que para todas las categorías de rendimiento académico la mayoría de los alumnos provienen de escuelas del ámbito Provincial y Municipal, pero con diferencias significativas en los porcentajes según rendimiento académico, alto: 78%, medio: 67%, bajo: 61%. Esto indica que el tipo de escuela secundaria que cursó el alumno está relacionado con el rendimiento académico logrado por el mismo, observándose que el porcentaje más alto de participación de escuelas del ámbito Provincial y Municipal (Estatal) corresponde a la categoría de mayor rendimiento académico.

- Considerando la cantidad de horas semanales que los alumnos dedicaron al estudio se observó que el 56% de quienes

han tenido un rendimiento académico alto han dedicado más de 20 hs semanales al estudio, ese porcentaje se reduce al 50% para el rendimiento académico medio y al 46% para el rendimiento académico bajo. Además, el 22% de quienes han tenido un rendimiento académico alto han dedicado hasta 10 hs semanales al estudio, ese porcentaje aumenta al 27% para el rendimiento académico bajo. Esto indica una relación directa entre la dedicación al estudio y el éxito académico.

· Considerando la importancia que los alumnos otorgan al estudio se observó que el 89% de quienes han tenido un rendimiento académico alto han otorgado más importancia al estudio que a la diversión, ese porcentaje se reduce al 50% para el rendimiento académico medio y al 64% para el rendimiento académico bajo, siendo del 64,6% para el total de la población. Además, el 11% de quienes han tenido un rendimiento académico alto han otorgado más importancia al estudio que al trabajo, ese porcentaje aumenta al 33% para el rendimiento académico medio y al 18% para el rendimiento académico bajo. Esto indica una relación entre el éxito académico y la importancia dada al estudio ante la diversión y el trabajo.

· Considerando los últimos estudios de la madre (los de mayor nivel) se observó que el 22% de quienes han tenido un rendimiento académico alto tienen madres con estudios de postgrado, ese porcentaje se reduce al 7% para el rendimiento académico bajo, siendo del 7,08% para el total de la población. Además, el 33% de quienes han tenido un rendimiento académico alto son hijos de madres con estudios universitarios completos, ese porcentaje decrece al 25% para el rendimiento académico medio y al 17% para el rendimiento académico bajo. Esto indica una relación entre el éxito académico y el nivel de estudios alcanzado por la madre.

· Considerando los últimos estudios del padre (los de mayor nivel) se observó que el 11% de quienes han tenido un rendimiento académico alto tienen padres con estudios de postgrado, ese porcentaje se reduce al 1% para el rendimiento académico bajo, siendo del 1,77% para el total de la población. Además, el 44% de quienes han tenido un rendimiento académico alto son hijos de padres con estudios universitarios completos, ese porcentaje decrece al 25% para el rendimiento académico medio y al 21% para el rendimiento académico bajo. Esto indica una relación entre el éxito académico y el nivel de estudios alcanzado por el padre.

· Teniendo en cuenta la opinión de los alumnos respecto de la utilización de las TIC se observó que el 56% de quienes han tenido un rendimiento académico alto consideraron que las TIC facilitan el proceso de enseñanza, ese porcentaje se reduce al 50% para el rendimiento académico medio, siendo del 53% para el bajo. Además, el 33% de quienes han tenido un rendimiento académico alto consideraron que será imprescindible el dominio de las TIC para el ejercicio profesional, ese porcentaje aumenta al 42% para los rendimientos académicos medio y bajo. Esto estaría indicando que la mayoría de los alumnos con mayor rendimiento académico estarían concentrados más en el proceso de enseñanza – aprendizaje que en el posible ejercicio futuro de la profesión.

Lo indicado precedentemente permite afirmar que el éxito y el fracaso académicos están relacionados con el tipo de escuela secundaria en la que cursó el alumno, la dedicación del alumno medida en horas semanales de estudio, la importancia otorgada al estudio frente a la diversión y al trabajo y el nivel de estudios de los padres y madres (La Red Martínez et al, 2015b, 2015c).

Objetivo general del nuevo proyecto

Obtener el conocimiento necesario para desarrollar un modelo predictivo para determinar rendimientos académicos potencialmente problemáticos, a partir de la implementación de técnicas de Minería de Datos.

Objetivos específicos del nuevo proyecto

- Utilizar técnicas de Minería de Datos sobre el Almacén de Datos construido con los datos de alumnos relevados durante 2013, 2014 y 2015 y extender el relevamiento de los datos de alumnos durante 2016 y 2017 para obtener conocimiento referido a las variables que predominantemente influyen en el rendimiento académico.

- Analizar y observar el comportamiento del modelo bajo condiciones típicas de rendimiento académico, obteniendo el conocimiento necesario para depurar el modelo predictivo.

- Validar el modelo mediante el seguimiento del rendimiento académico de estudiantes-tipo seleccionados, obteniendo un modelo predictivo confiable.

- Utilizar el modelo predictivo de rendimiento académico desarrollado para predecir la probabilidad de que un estudiante abandone el cursado de una asignatura, dadas sus características socioeconómicas y académicas.

Formación de Recursos Humanos

El equipo de trabajo está integrado por un Director (Doctor, Categoría II P.I., Categoría A UTN), un Co-director (Doctor, Categoría IV P.I., Categoría C UTN), un investigador (Especialista) realizando su tesis de maestría, un investigador (Magister) y dos becarios. Actualmente se está trabajando en la definición de planes de tesis de maestría con temáticas afines a la del proyecto.

Referencias

Briand, L. C., Daly, J., y Wüst, J. (1999) A unified framework for coupling measurement in object oriented systems, *IEEE Transactions on Software Engineering*, 25, 1.

Connolly, T. y Begg, C. (2005) *Sistemas de bases de datos. Un enfoque práctico para diseño, implementación y gestión*. 4ta Edición. Ed. Pearson Addison Wesley.

La Red Martínez, D.; Karanik, M.; Giovannini, M. Y Pinto, N. (2015) *Academic Performance Profiles: A Descriptive Model Based on Data Mining*; Volume 11 – N° 9; *European Scientific Journal (ESJ)*; pp. 17-38; ISSN N° 1857-7881; University of the Azores, Portugal.

La Red Martínez, D. L.; Podestá, C. E. (2014a) *Data Mining to Find Profiles of Students*; Volume 10 – N° 30; *European Scientific Journal (ESJ)*; pp. 23-43; ISSN N° 1857-7881; University of the Azores, Portugal.

La Red Martínez, D. L.; Podestá, C. E. (2014b) *Contributions from Data Mining to Study Academic Performance of Students of a Tertiary Institute*; Volume 02 – N° 9; *American Journal of Educational Research*; pp. 713-726; ISSN N° 2327-6126; U.S.A.

Maletic, J. I., Collard, M. L., y Marcus, A. (2002) *Source Code Files as Structured Documents*, in *Proceedings 10th IEEE International Workshop on Program Comprehension (IWPC'02)*, Paris, France, pp. 289-292.

Maradona, G. y Calderón, M. I. (2007) *Una aplicación del enfoque de la función de producción en educación*. *Revista de Economía y Estadística*, Universidad Nacional de Córdoba, XLII. Argentina.

Marcus, A. (2003) *Semantic Driven Program Analysis*. Kent State University, Kent, OH, USA, Doctoral Thesis.

Porto, A. y Di Gresia, L. (2003) *Características y rendimiento de estudiantes universitarios. El caso de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata*. Documentos de Trabajo, Universidad Nacional de La Plata.

Reyes R, S. L. (2004) *El Bajo Rendimiento Académico de los Estudiantes Universitarios. Una Aproximación a sus Causas*. *Revista Theoretikos*. Año VI, N° 18, enero-junio, El Salvador.

Wilson, R. L.; Hardgrave, B. C. (1995) *Predicting graduate student success in an MBA program: Regression versus classification*. *Educational and Psychological Measurement*, 55, 186-195. USA.

TIC y Objetos de Aprendizaje en el Ámbito Educativo

Zulema Beatriz Rosanigo⁽¹⁾, Pedro Bramati⁽²⁾, Claudia López de Munain⁽¹⁾, Silvina Bramati⁽²⁾,
Leda Cotti de La Lastra⁽²⁾

brosanigo@yahoo.com.ar, pedrobramati@speedy.com.ar, klaucvj@ing.unp.edu.ar,
silvina.bramati@gmail.com, ledacotti@hotmail.com

⁽¹⁾ Departamento de Informática – ⁽²⁾ Departamento de Ingeniería

Facultad de Ingeniería – Sede Trelew – U.N.P.S.J.B. – Te-Fax (0280) 4428402

RESUMEN

En el presente trabajo se presentan objetivos y principales resultados de la línea de investigación sobre TIC y Objetos de Aprendizaje (OA) que se lleva a cabo en la Facultad de Ingeniería (Sede Trelew) de la Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB).

Los principales ejes abordados son: Uso innovador de tecnología en ambientes educativos, Modelos y escenarios para la integración de objetos de aprendizaje, Metodología y patrones de diseño de OA, consideraciones para la evaluación de los contenidos, Análisis de herramientas de diseño de material educativo, Análisis de dinámicas de interacciones, Formación basada en competencias, Recursos y estrategias para un aprendizaje activo del alumno.

El proyecto contribuye a la inserción de la cultura digital, fortaleciendo su empleo en actividades de docencia, de investigación y de comunicación interna y externa de la comunidad universitaria.

Palabras clave: TIC, Objeto de aprendizaje, ambientes virtuales de enseñanza aprendizaje.

CONTEXTO

El contexto de aplicación del proyecto “TIC y Objetos de Aprendizaje en ambientes multidisciplinares de investigación y enseñanza de la ingeniería” se enmarca dentro de la Facultad de Ingeniería (Sede Trelew) de la UNPSJB. Este proyecto está aprobado y financiado por la Secretaría de Ciencia y Técnica de la UNPSJB. Se inició en Enero de 2014 y su fecha de finalización es en Diciembre de 2016.

Intervienen docentes e investigadores provenientes de distintas disciplinas y alumnos de las carreras de Ingeniería Civil y Licenciatura en Informática/ Sistemas.

1. INTRODUCCIÓN

Una de las funciones más importantes que tiene el docente del siglo XXI es lograr que los estudiantes adquieran hábitos óptimos para responder adecuada y eficientemente a las necesidades de este milenio.

La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) aplica una estrategia amplia e integradora en lo tocante a la promoción de las TIC en la educación, que procura lograr una mayor comprensión de cómo las

tecnologías pueden ser utilizadas para mejorar el acceso, la equidad y la calidad de la educación en todo el mundo [1, 2, 3].

Según Rivoir, Pittaluga, Di Landri, Baldizán & Escuder [4], las TIC *“atravesan tanto procesos macro-sociales -como su rol en la consolidación del proceso de globalización, el aumento de la importancia de la información y el conocimiento en los procesos productivos- hasta cambios producidos a nivel cultural y simbólico, así como transformaciones de la vida cotidiana de las personas”*.

Aprovechar adecuadamente las tecnologías de información y comunicaciones (TIC) en el ámbito educativo es crucial para facilitar a los educadores de las herramientas necesarias para impactar creativamente el proceso de enseñanza-aprendizaje y superar los retos y desafíos [5].

A partir de la mayor interacción que se genera en Internet, los docentes tienen la posibilidad de interactuar directamente con contenidos, creando y compartiendo con sus pares dándole una dimensión más *“social y participativa”* [6]. Los propios alumnos empiezan a manejar distintos aspectos de la experiencia de aprendizaje, incluida la creación y el aprovechamiento compartido de conocimientos fundamentales [7].

En este contexto el aprendizaje dejó de ser concebido como un proceso individual para convertirse en un proceso constructivo, participativo y social en el que se realizan actividades conjuntas, apoyándose en la tecnología, siendo el aprendizaje el resultado de la relación interactiva entre profesor, alumno y contenidos. El alumno se convierte en protagonista de su propio proceso de aprendizaje y él mismo adquiere contenidos, destrezas y habilidades. [8]

Los ambientes de aprendizaje se han flexibilizado en tiempo y espacio mediante las herramientas de comunicación y colaboración, transformándose en espacios de interacción virtual, donde se fortalecen las relaciones entre los profesores, los

alumnos y los contenidos a enseñar. En estos esquemas, la actividad del estudiante se considera como agente, protagonista principal y responsable último de su aprendizaje [9]. Y a través de la combinación de múltiples medios y acompañados de un buen diseño instruccional, se pueden ofrecer nuevas alternativas pedagógicas que incluyan la diversidad [10, 11].

A diferencia del proceso de enseñanza aprendizaje tradicional, en el que el profesor marca el ritmo y dirige la actividad, en la enseñanza apoyada en las TIC, el rol del docente se ve multiplicado y acompaña al alumno en su proceso de aprendizaje, actuando de organizador y facilitador de la participación [8]. Le proporciona instrumentos de acceso al medio, de desarrollo, de construcción y de exploración de múltiples perspectivas, favoreciendo así su inmersión en un contexto para el aprendizaje activo. En la medida que se intensifica el involucramiento del alumno, se maximiza el aprovechamiento y retención de los aprendizajes.

Este cambio educativo hacia un modelo centrado en el alumno ha propiciado el desarrollo de los objetos de aprendizaje (OA).

Un OA es un conjunto de recursos, autocontenible, diseñado y creado en pequeñas unidades *digitales*, con un *propósito educativo* para maximizar el número de situaciones en las que se puede utilizar [12, 13, 14, 15, 16].

Los OA permiten compartir y reutilizar recursos educativos en procesos de aprendizaje apoyados por tecnología. Ofrecen la posibilidad de tener contenidos educativos reutilizables, autocontenidos, independientes de la plataforma de uso, y además permiten elaborar múltiples y flexibles itinerarios pedagógicos que se adaptan a las necesidades específicas de los alumnos, fortaleciendo la educación. Permiten incluir una alta interactividad para apoyar el aprendizaje activo,

personalizarlos para ciertos usuarios, considerando estilos de aprendizaje, accesibilidad, realidad aumentada, dispositivos móviles y ubicuidad [17, 18, 19, 20].

La utilización de las TIC en el mundo educativo es aún muy inferior a la evolución en otros ámbitos tales como el mundo empresarial y de negocios [19]. Indudablemente, su incorporación en ambientes educativos requiere de nuevas competencias docentes que sólo se logran a través de la capacitación y de una profunda reflexión acerca de su potencial utilidad para promover procesos de construcción de conocimientos científicos [21, 22]. Las habilidades de un docente en el siglo XXI anidan en competencias instrumentales informáticas, competencias comunicacionales, competencias sociales, competencias de uso didáctico de la tecnología, competencias para la docencia virtual [23].

La naturaleza integral del concepto competencia educativa posibilita la concreción de los cuatro pilares de la educación del siglo XXI: aprender a conocer, aprender a hacer, aprender a vivir juntos y aprender a ser [24]. Supone saber hacer y aplicar conocimiento, más que adquirirlo de forma receptiva. Implica nuevas metodologías y formas de evaluación, supone reflexionar y ampliar el concepto de alfabetización digital, los modelos de formación del profesional de la enseñanza y aunar las teorías del aprendizaje de pedagogos de la escuela activa liderada por Piaget, del constructivismo social de Vigotsky y de la socialización del aprendizaje de Freire [21].

Actualmente se reconoce que las competencias propician un mayor acercamiento entre los conocimientos y el desempeño, y persiste una demanda social hacia la formación de profesionales competentes, capaces de adaptarse a los nuevos requerimientos laborales, sociales y tecnológicos, para responder positivamente a situaciones específicas y tomar decisiones

que les permitan resolver problemas en forma eficaz y eficiente [25].

Con este proyecto se pretende integrar y fomentar el empleo de TIC en situaciones de enseñanza/aprendizaje y de investigación. Para lo cual se plantean los siguientes objetivos particulares:

- a) Explorar sobre procesos y herramientas para el diseño de objetos de aprendizaje (OA) e integración de TIC en docencia e investigación.
- b) Identificar prácticas y experiencias educativas y de investigación relacionadas con la utilización de las TIC, así como los modelos organizativos que favorecen dichas situaciones.
- c) Promover y favorecer el desarrollo y utilización de TIC, y particularmente OA, en la comunidad educativa.
- d) Implementar soluciones concretas a problemas de educación apoyada en tecnología.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Se viene trabajando en la línea de Informática Educativa desde 1997.

El proyecto actual, “TIC y Objetos de Aprendizaje en ambientes multidisciplinares de investigación y enseñanza de la ingeniería”, iniciado en 2014 procura integrar y fomentar el empleo de TIC y OA en situaciones de enseñanza/aprendizaje y de investigación.

Los principales ejes abordados son:

- Integración de TIC en el proceso de enseñanza-aprendizaje. Uso innovador de tecnología en ambientes educativos, experiencias, uso de tecnología móvil,

trabajo colaborativo mediado por TIC. El objetivo es relevar, analizar y evaluar las experiencias que se desarrollan en el espacio virtual de aprendizaje. Desde la perspectiva del alumno, se busca identificar indicadores de satisfacción y del impacto sobre los aprendizajes. Desde la perspectiva del docente, se determinan las necesidades formativas y se promueven instancias de capacitación y experimentación.

- Modelos y escenarios para la integración de objetos de aprendizaje y herramientas de gestión de contenido.
- Metodología y patrones de diseño de OA, criterios de evaluación de la calidad del OA, consideraciones para la evaluación de los contenidos.
- Análisis de herramientas de diseño de material educativo. Experimentación con herramientas de software libre que facilitan la generación de materiales educativos, especialmente aquellas que permitan crear actividades interactivas y puedan ser usadas por docentes con pocos conocimientos informáticos.
- Análisis de dinámicas de interacciones suscitadas en el EVEA (entorno virtual de enseñanza-aprendizaje) en el marco de las actividades educativas de un grupo de estudiantes de curso de posgrado.
- Formación basada en competencias. Análisis de competencias digitales del profesorado y del alumnado.
- Formación de recursos humanos en el uso de TIC.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se han analizado diversas experiencias en la implementación educativa de OA y uso

innovador de la tecnología en ambientes educativos, profundizando en la aplicación de plataformas virtuales de aprendizaje a través de dispositivos móviles (Tablet pc y smartphones). Se han estudiado herramientas específicas para llevar adelante aprendizaje y trabajo colaborativo en procesos educativos y se ha desarrollado una metodología de trabajo colaborativo mediada por TIC.

Se ha incursionado en la utilización de recursos de Realidad Aumentada (RA). La RA es una tecnología con capacidad de enriquecer elementos de la realidad con información detallada pudiendo aplicarse en ámbitos muy variados. En el ámbito educativo se utiliza, entre otras cosas, para complementar los materiales didácticos con modelos virtuales que estimulen la percepción y ayuden a la comprensión de los conceptos.

Se ha avanzado en el desarrollo de un componente de software para generar dispositivos interactivos de autoevaluación para incluir en los OA.

Se han estudiado las interacciones que se producen en el EVEA con la finalidad de comprender cómo se produce el proceso de enseñanza y aprendizaje en un espacio virtual colaborativo y a partir de ello, realizar correcciones que permitan mejorar el rendimiento académico de los estudiantes. Se están teniendo en cuenta los estilos de aprendizaje y cómo inciden en el proceso enseñanza y aprendizaje. Con base en el análisis cuantitativo y cualitativo de la información obtenida de la plataforma utilizada para impartir cursos de formación en OA, se detectaron patrones de conducta, permitiendo establecer relaciones existentes entre los participantes, que sirven para identificar grupos de comportamiento homogéneo, para detectar fortalezas y debilidades en el desarrollo de los cursos, tanto en el material utilizado, las tutorías como así también en el uso del aula virtual.

Con el objetivo de contribuir a la inserción de la cultura digital, fortaleciendo su empleo en actividades de docencia, de investigación y de comunicación interna y externa de la comunidad universitaria, se están realizando capacitaciones para docentes de distintos niveles educativos, y así formar a una gran masa de docentes en la producción de objetos de aprendizaje utilizando TIC y herramientas de la web 2.0.

4. FORMACIÓN DE RECURSOS HUMANOS

Se busca fortalecer y afianzar el trabajo interdisciplinario y el establecimiento de líneas de investigación en estas temáticas.

Dos integrantes han obtenido el título de Magister. Un integrante está realizando el doctorado. Se ha contribuido a la concreción de dos tesinas de grado en el área de informática. Dos alumnos están desarrollando sus tesinas.

Se continúa con capacitación en diseño y producción de OA, tanto para docentes de nivel universitario como nivel medio e inicial. Estos cursos ofrecen un espacio de formación y acompañamiento para el diseño y creación de materiales educativos en espacios virtuales. Los cursantes aprenden a diseñar y crear objetos de aprendizaje interoperables, reutilizables, compartibles para contenidos educativos en un contexto crítico y reflexivo acerca del uso de esta tecnología en educación.

5. BIBLIOGRAFÍA

1. ANDERSON, J. (2010), *ICT Transforming Education - A Regional Guide*, UNESCO Bangkok, ISBN 978-92-9223-326-6.
2. SCHALK QUINTANAR, A. M. (2010),

El impacto de las TIC en educación, OREALC UNESCO.

3. WEST, M. (2012) *Aprendizaje móvil para docentes*, UNESCO 2012, ISSN 2305-8617.
4. RIVOIR, A., PITTALUGA, L., DI LANDRI, F., BALDIZÁN, S., & ESCUDER, S. (2011). *El Plan Ceibal: Impacto comunitario e inclusión social. 2009-2010*. Recuperado de <http://www.observatic.edu.uy/wp-content/uploads/2011/04/Informe-Final-CEIBAL-inclusi%C3%B3n-social-Rivoir-Pittaluga.pdf>
5. UNESCO (2009). Observatory Portal: Monitoring the Development of the Information Society towards Knowledge Societies, Communication and Information. Recuperado de: <http://www.unesco.org/webworld/observatory/>
6. COBO, R. Y PARDO K. (2007). *Planeta Web 2.0. Inteligencia colectiva o medios fast food*. Barcelona: Grup de Recerca d'Interaccions Digitals, FLACSO México.
7. COBB, J. T. (2008). *Learning 2.0 for Associations*. Recuperado de: http://blog.missiontolearn.com/files/Learning_20_for_Associations_eBook_v1.pdf
8. CABERO, J. (2007). *Tecnología Educativa*. Ed. Mac Graw Hill.
9. PÉREZ DE A., MARÍA DEL C.; TELLERIA, MARÍA B. (2012) *Las TIC en la educación: nuevos ambientes de aprendizaje para la interacción educativa*. *Revista de Teoría y Didáctica de las Ciencias Sociales*, núm. 18, enero-diciembre, 2012, pp. 83-112 Universidad de los Andes
10. DICK, W., CAREY, L. Y CAREY, J. (2005). *The systematic design of instruction*, (6th ed.). USA: Person.
11. GAGNE, R, BRIGGS, L & WAGNER,

- W (1992). Principios de diseño instruccional (4^a ed.), Holt, Reihhart, y Winston Inc.
12. WILEY, David A. (2001) "Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy". <http://www.elearning-reviews.org/topics/technology/learning-objects/2001-wiley-learning-objects-instructional-design-theory.pdf>
 13. CHIAPPE, A. (2009). Objetos de aprendizaje 2.0: una vía alternativa para la reproducción colaborativa de contenido educativo abierto. Colección: Univirtual Objetos de Aprendizaje. Prácticas y perspectivas educativas ISBN: 958-8162-65-3 Pontificia Universidad Javeriana – Cali.
 14. POLSANI (2003). Use and abuse of reusable Learning Objects Pithamber R. Journal of Digital Information, Volume 3 Issue 4 Article No. 164, 2003-02-19
 15. ROSANIGO, Z. B. (2013) Objetos de Aprendizaje en "Capacitación y Gestión del Conocimiento a través de la Web 2.0" DYKINSON S.L. Madrid.
 16. CHIAPPE A. (2012) Resultados de Investigación- Objetos de Aprendizaje Móviles- Transposición Didáctica. <http://www.scivee.tv/node/39529> (consulta 12-05-2013).
 17. CORONA FLORES, J.D. y GONZÁLEZ BECERRA, B.L. (2012). Objetos de aprendizaje: Una Investigación Bibliográfica y Compilación. *RED, Revista de Educación a Distancia*. Número 34. 15 de noviembre de 2012. Consultado el 10/07/2013 en <http://www.um.es/ead/red/34>
 18. CRUZ MELÉNDEZ, A.; ALFARO RIVERA, J. A.; RAMÍREZ MONTROYA, M. S. (2012). Objeto de aprendizaje abierto para la formación docente orientado a desarrollar competencias de pensamiento crítico con énfasis en habilidades cognitivas. *RIED. Revista Iberoamericana de Educación a Distancia*, volumen 15, nº 1, pp. 103-125.
 19. WATSON, D. (2001) Pedagogy before technology: re-thinking the relationship between ICT and teaching. *Education and Information technologies*. V. 6, Issue 4, P. 251 - 266.
 20. KAY, R.; KNAACK, L. (2009). Assessing learning, quality and engagement in learning objects: the Learning object evaluation scale for students (LOES-S). *Educational Technology, Research and Development*, 57 (2), (147). ProQuest Academic Research Library.
 21. AREA, M., GUTIÉRREZ, A., Y FERNÁNDEZ, V. (2012). Alfabetización digital y competencias informacionales. Barcelona, España. Ariel.
 22. TAPIA, H. (2012). Las Tecnologías de la Información y la Comunicación en la Formación Inicial Docente. Contexto, estándares, metodología y actitudes hacia las TIC en el proceso de formación. Tesis Doctoral. Universidad de Navarra, Pamplona, España.
 23. MARTIN CUADRADO, A. M (2015). *El desarrollo de las competencias tic en los contextos formativos: entrenamiento para la integración social*. (cap.16) en Gallego G., Alvarez M., Rosanigo, Z.B. y Cela K. (Coordinadores) *TIC y Web 2.0 para la inclusión social y el desarrollo sostenible*. Madrid Ed. Dykinson S.L. ISBN 978-84-9031-412-8
 24. DELORS, J. (1999). La educación encierra un tesoro. España. Santillana.
 25. Rychen, D. y Salganik, L. (Eds.) (2001) *Defining and selecting key competencies*, París: Organization for Economic Cooperation and Development.

Aplicaciones para la enseñanza aprendizaje de matemática en el nivel secundario

Ascheri María Eva, Testa Oscar, Pizarro Rubén, Camiletti Pablo, Díaz Lucas,
Dimartino Santiago

Departamento de Matemática / Facultad de Ciencias Exactas y Naturales/
Universidad Nacional de La Pampa
Uruguay 151, 00-54-02954-425166
mavacheri@gmail.com ; otesta@gmail.com

Resumen

Luego de una primera etapa del proyecto se ha realizado un relevamiento de la cantidad de dispositivos móviles con Android que se encuentra presente en las instituciones educativas de nivel medio de la ciudad de Santa Rosa, La Pampa. Posteriormente se realizó un análisis de las aplicaciones existentes, para dispositivos móviles destinadas a la enseñanza de contenidos de Matemática.

Se comenzó a partir de estos datos el diseño de aplicaciones educativas con las que se experimentó en las aulas. Con los resultados obtenidos se realizarán los ajustes necesarios en el diseño de las aplicaciones. Para dar respuestas a esta tarea y a la complejidad que implica determinar las variables que se deberán considerar, se abordará una metodología de investigación y desarrollo de tipo cualitativa, utilizando un diseño de estudio de casos, con técnicas de observación participante, cuestionarios abiertos y entrevistas.

Palabras clave: dispositivos móviles, android, enseñanza aprendizaje.

Contexto

El proyecto de investigación que se presenta está radicado en el Departamento

de Matemática de la Facultad de Ciencias Exactas y Naturales de la UNLPam y financiado por dicha Facultad.

Tiene una duración de tres años, siendo este el tercer año de su desarrollo. Está vinculado con otros proyectos del Departamento ya ejecutados y en ejecución, relacionados con la enseñanza de la matemática y la inclusión de tecnologías en educación.

Introducción

Existen actualmente gran cantidad de aplicaciones desarrolladas para equipos con sistema operativo Android y destinadas especialmente a educación (Ralph, 2011).

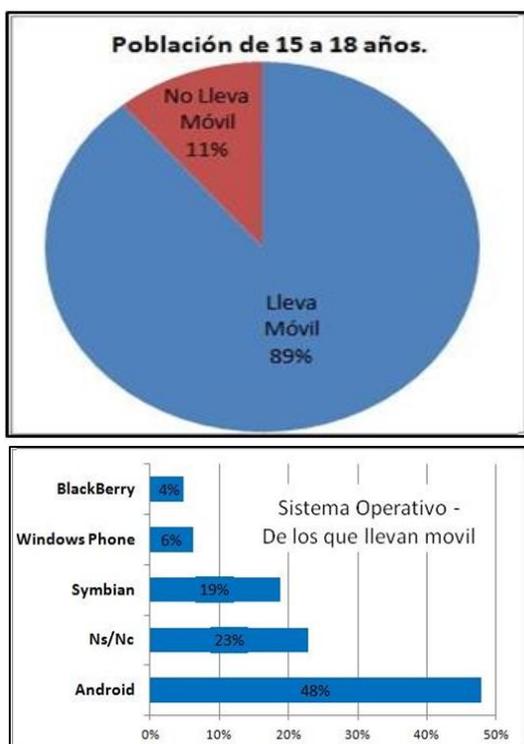
Sin embargo, a pesar de varios años de trabajo en el campo del aprendizaje móvil –explican en el *Infokit de aprendizaje móvil* de JISC InfoNet–, el cuerpo de investigación disponible es relativamente pequeño. “Hay dos razones para ello. En primer lugar, la rápida evolución de los dispositivos móviles ha causado problemas para el trabajo longitudinal significativo. A menudo, cuando los dispositivos adquiridos institucionalmente comienzan a ganar tracción pueden ser rechazados por ser obsoletos. En segundo lugar, concepciones culturales han impedido el

uso de dispositivos móviles en las instituciones educativas. Vistos como perjudiciales, que distraen la atención o que causan problemas de privacidad, la política de gestión en muchos de esos ámbitos ha sido de prohibición absoluta”.

Presencia de móviles entre estudiantes y docentes.

¿Cuál es el nivel actual de presencia de dispositivos móviles entre los estudiantes y docentes de los colegios de nivel secundario de la ciudad de Santa Rosa?

Nos encontramos que en la franja etaria que va de 15 a 18 años (30% de la muestra de 166 encuestas practicadas en la ciudad de Santa Rosa, La Pampa), el 89% de los estudiantes lleva el móvil a la escuela. De estos el 48% afirma que su móvil cuenta con sistema Android, en tanto que un 23% no sabe qué sistema posee.



Gráficos obtenidos a partir de 166 encuestas realizadas en colegios secundarios de la ciudad de Santa Rosa, La Pampa (Argentina)

En esta franja etaria nos encontramos con que la mitad de los estudiantes encuestados posee sistema Android y existe casi una cuarta parte que podría poseerlo y no lo sabe.

La utilización de dispositivos móviles favorece la adquisición de un conjunto de competencia en diferentes áreas del conocimiento, entre ellas, matemática. En este caso, favorece la capacidad para utilizar y relacionar números, sus operaciones básicas y el razonamiento matemático y la capacidad para interpretar la información, ampliar conocimientos y resolver problemas tanto de la vida cotidiana como del mundo laboral.

Existe una gran variedad de aplicaciones para móviles de calculadoras científicas que, combinadas con una interfaz táctil capaz de mostrar cualquier tipo de botones y una pantalla gráfica, tienen grandes posibilidades educativas (Valero et al, 2012).

Ante el surgimiento de estos nuevos dispositivos y las herramientas informáticas a las que permiten acceder, como profesionales que desarrollamos nuestra actividad en la formación, nos proponemos investigar la forma de utilizar aplicaciones educativas existentes o de desarrollar nuevas aplicaciones para que se puedan incorporar al proceso de enseñanza – aprendizaje de la matemática utilizando dispositivos móviles. Debemos entonces proponer nuevas alternativas para el desarrollo de los contenidos en el nivel secundario, incorporando las aplicaciones educativas para dispositivos móviles.

Pero no sólo debemos adoptar e incorporar aplicaciones ya existentes diseñadas en ámbitos ajenos al nuestro y al de nuestros estudiantes de nivel medio. Los integrantes del proyecto poseen los conocimientos sobre temas informáticos y pedagógicos que les permitirán dar un marco y desarrollar estas tareas. Para ello,

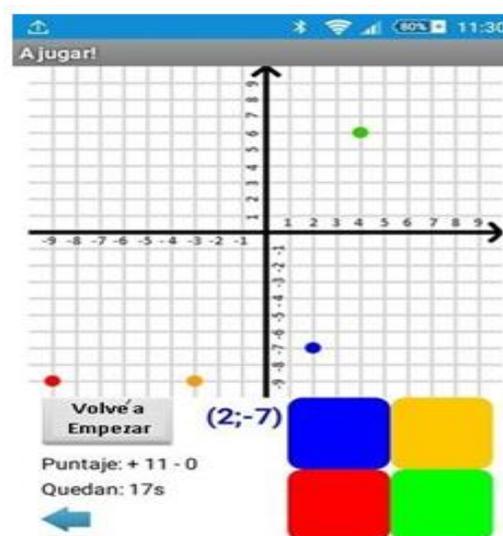
se seguirán las etapas de diseño, experimentación y ajustes al diseño.

En este momento del proyecto estamos en la búsqueda de alternativas para el desarrollo de contenidos y formas de apropiación de los procedimientos / conocimientos matemáticos.

Las aplicaciones desarrolladas deben reúnen al menos algunas de las características de la mecánica del juego (gamificación) o ludificación según Cortizo Pérez, J.C. et al (2011), entre otras: metas y reglas definidas, sistema de retroalimentación o feedback y motivación. Por ejemplo una característica distintiva es la de lograr aplicaciones que den retroalimentación inmediata a preguntas al evaluar las respuestas practicada por el usuario.

Buscamos promover en las aplicaciones un desafío accesible acorde a distintos niveles a los que se pueda acceder a medida que se van superando los mismos, que estas pueda involucrar al usuario/alumno ya sea porque le permite sumar puntos, avanzar a distintos escenarios o completar misiones. Es decir al mismo tiempo que mejora su comprensión de conceptos importantes para la asignatura, es recompensado por los niveles y el reconocimiento que estos representan Estanyol, E. et al (2013)

Una de las aplicaciones desarrolladas propone reconocer pares ordenados de puntos en el plano, se muestra un plano con puntos de distintos colores y al mismo tiempo se muestra un par ordenado, el jugador tiene que elegir el color del punto al que corresponde el par ordenado. Con similares características planificamos desarrollar nuevos niveles que impliquen reconocer pendientes y ordenadas al origen, ecuaciones completas y hasta la ecuación que corresponda a situaciones planteadas con similares formas de registrar las respuestas del usuario en pantallas similares.



Pantalla de la aplicación desarrollada para dispositivos con Android que propone identificar pares ordenados, reconocer funciones y ecuaciones

La herramienta seleccionada en este caso para el desarrollo de aplicaciones en Android es App Inventor, una herramienta visual de arrastrar y soltar. Como indican Wolber et al (2011), el programador puede diseñar la interfaz de usuario (el aspecto visual) de una aplicación utilizando una interfaz gráfica basada en web, y a continuación especificar el comportamiento de la aplicación “juntando bloques” como si estuviera armando un rompecabezas. Muchas son las ventajas de este tipo de programación; dentro de las más destacadas, se pueden mencionar: la agilidad de desarrollo, el tiempo muy corto de armado de nuevas aplicaciones, el entorno gráfico y amigable que abre las puertas para que personas con muy pocos conocimientos de programación puedan desarrollar aplicaciones que le sean de utilidad en el día a día.

A partir del marco teórico abordado y conociendo que actualmente en el ámbito de aplicación del proyecto la utilización de dispositivos móviles es prácticamente nula (por las razones que se indicaron en el

segundo párrafo, JISC Advance), hemos planificado llegar a diferentes instituciones educativas y trabajar en las aulas.

El proyecto presenta como principal innovación, no sólo realizar un relevamiento y diseño de aplicaciones educativas, sino que además, se espera acompañar el proceso de enseñanza aprendizaje en el cual se incluyan las mismas.

Líneas de Investigación, Desarrollo e Innovación

El paso inmediato del proyecto es utilizar las aplicaciones que hemos desarrollado hasta el momento en escuelas de educación secundaria, en esta prueba de campo observaremos las variables didácticas en interacción, recuperando beneficios y dificultades, el contexto institucional y el status asignado a la experiencia, las formas de registro y seguimiento, la interacción, de los estudiantes entre ellos y con sus docentes, las secuencias didácticas, los usos en el contexto escolar y los posibles usos alternativos de las aplicaciones.

Diferencial que podremos observar en tanto la herramienta cumpla con funciones pragmáticas, epistémicas y/o heurística que podamos observar en las formas de comprender, actuar y transformar el entorno, como proceder, organizar y controlar acciones y procedimientos.

Todos los datos y percepciones que podamos registrar serán analizados e inmediatamente nos servirán para ensayar nuevos prototipos que descarten los inconvenientes, potencien los beneficios y tengan en cuenta también los aspectos valorados por docentes y estudiantes de la asignatura matemática en el nivel secundario.

Las líneas de investigación que se abordan con el presente proyecto son las siguientes:

- Nivel de penetración de los dispositivos móviles en las actividades de enseñanza aprendizaje en aulas de nivel medio.
- Búsqueda, diseño y desarrollo de aplicaciones educativas para la enseñanza aprendizaje de matemática en el nivel medio, utilizando la herramienta APP Inventor
- Características metodológicas a implementar en el desarrollo de clases incluyendo dispositivos móviles.

Resultados y Objetivos

Los objetivos planteados son los siguientes:

- Conocer, analizar y difundir las características de las aplicaciones, para dispositivos con sistema operativo Android, destinadas a la enseñanza - aprendizaje de matemática.
- Desarrollar nuevas aplicaciones, para dispositivos móviles con sistemas operativos Android, para la enseñanza – aprendizaje de matemática.
- Experimentar, en distintos centros educativos de nivel secundario, las aplicaciones educativas en los dispositivos móviles.
- Capacitar a docentes de nivel medio en la utilización y desarrollo de aplicaciones para dispositivos móviles con sistema operativo Android.

Para dar cumplimiento a los objetivos planteados se ha comenzado por diseñar e implementar un relevamiento de información en diferentes cursos de nivel medio. Esta actividad, tendiente a conocer la presencia de dispositivos móviles con sistema operativo Android en dicha

población, arrojó importantes resultados que permitió confirmar la amplia presencia de estos dispositivos.

Algunos análisis preliminares relacionados al desarrollo de este proyecto se han presentado en distintos eventos (Ascheri et al, 2014 a y b).

Formación de Recursos Humanos

El equipo de trabajo es de carácter interdisciplinario, está conformado con especialistas del área de Educación, Matemática y Computación. Participan además dos estudiantes del Profesorado en Computación.

Se realizaron, también, actividades de capacitación y transferencia referidas a la utilización de aplicaciones educativas para dispositivos móviles con sistema operativo Android.

Referencias

- **Ascheri, M., Testa, O., Pizarro, R., Camiletti, P, Díaz L.,** “Utilización de dispositivos móviles con sistema operativo Android para matemáticas. Una revisión de aplicaciones”, V REPEM, ISSN N° 2362-5716, Vol. 5, pp. 287-292. (2014). Disponible en <http://repep.exactas.unlpam.edu.ar/descargas/Memorias%20de%20REPEM2014.pdf>
- **Cortizo Pérez, J., Carrero García, F. Monsalve Piqueras, B. Velasco Collado, A., Díaz Del Dedo, L., Pérez Martín, J** (2011). Gamificación y Docencia: Lo que la Universidad tiene que aprender de los Videojuegos. VIII Jornadas Internacionales de Innovación Universitaria, ed. Disponible en http://www.aulavirtual.urjc.es/moodle/pluginfile.php/2103055/mod_resource/content/1/46_Gamificacion.pdf fecha último acceso 27/07/ 2015.
- **Estanyol, E. Montaña, M. Lalueza, F.** (2013) “Comunicar jugando. Gamification en publicidad y relaciones públicas” (pag.111-120) International Conference On Communication And Reality: Breaking The Media Value Chain disponible en: http://www.researchgate.net/publication/237841789_Comunicar_jugando_Gamification_en_publicidad_y_relaciones_publicas Consultado el 27/07/ 2015 fecha último acceso 27/07/ 2015
- **JISC Advance** “Informe infoKit de aprendizaje móvil”. (2011). El informe puede descargarse completo en pdf desde la siguiente dirección: <https://files.pbworks.com/download/XvIzfbp3DN/mobilelearninginfokit/50757490/mobile-learning-infokit-spanish.pdf>
- **Ralph, R.** “Google App Inventor: Beginner's Guide” Packt Publishing, Primera Edición. (2011).
- **Valero, C., Riura Redondo, M., Sánchez Palacín, A.,** “Tendencias actuales en el uso de dispositivos móviles en educación”. Educational Portal of the Americas – Department of Human Development, Education and Culture © OEA-OAS ISSN 0013-1059 La Educación Digital Magazine N 147. (2012). Disponible en: http://educoas.org/portal/la_educacion_digital/147/pdf/ART_UNNED_EN.pdf
- **Wolber, D., Abelson, H., Spertus, E. & Looney, L.** “App Inventor. Create Your Own Android Apps”. O'Reilly, Primera Edición. (2011).

Software educativo para la resolución numérica y gráfica de ecuaciones diferenciales ordinarias

Ascheri M. E., Pizarro R., Astudillo G., García P., Culla M. E., Pauletti C.
Facultad de Ciencias Exactas y Naturales. Universidad Nacional de la Pampa
Avda. Uruguay 151 - Santa Rosa - La Pampa - 02954-425166
mavacheri@gmail.com rubenpizarro71@gmail.com

Resumen

En el marco del curso Cálculo Numérico que se dicta para varias carreras de la Facultad de Ciencias Exactas y Naturales de la UNLPam, se está trabajando en un proyecto de investigación sobre la elaboración de un software educativo para la *resolución numérica y gráfica de integrales y de ecuaciones diferenciales ordinarias*, usando herramientas gratuitas existentes en la Web.

Los objetivos que se pretenden alcanzar son:

- Generar el contexto educativo adecuado a los contenidos a desarrollar y a los objetivos propuestos en la asignatura.
- Proporcionar una herramienta tecnológica para afianzar los conceptos teóricos y la aplicación de los métodos numéricos a través de la visualización gráfica del funcionamiento de los mismos.
- Poner a disposición de los estudiantes una herramienta de acceso libre y disponible en la Web.

En una primera instancia, se comenzó a trabajar con la elaboración del software educativo que incluye el tema *resolución numérica y gráfica de integrales*. Se elaboró el software educativo correspondiente, se experimentó, analizó y validó. Esto es, se hicieron todas las pruebas estipuladas y los ajustes necesarios para su aplicación en el curso de Cálculo Numérico.

En el transcurso de este año, se comenzará a hacer lo propio abordando ahora el tema *resolución numérica y gráfica de ecuaciones diferenciales ordinarias*.

Palabras clave: software educativo, integración numérica, ecuaciones diferenciales ordinarias.

Contexto

El presente Proyecto de Investigación, está acreditado y financiado por la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa. Surge como continuidad de un proyecto ya finalizado relativo a la elaboración de un software educativo para la *resolución numérica de ecuaciones no lineales, interpolación y aproximación polinomial y ajuste de curvas por mínimos cuadrados*.

Está relacionado con otros proyectos del Departamento de Matemática del cual este grupo forma parte, en los que se abordan contenidos relacionados a la matemática, su enseñanza y la inclusión de tecnologías en el proceso de enseñanza y de aprendizaje.

Introducción

La experiencia obtenida a partir de la estrategia metodológica empleada en un Proyecto de Investigación anterior, motiva a que se continúe con el desarrollo de software educativo que incluya otras temáticas de Cálculo Numérico para lograr una integración curricular y complementar

los resultados obtenidos hasta el momento. Esta herramienta tiene como funcionalidad pretendida, guiar el aprendizaje como soporte del proceso de enseñanza. Actualmente, se ha finalizado con la etapa que se refiere a la elaboración del software educativo que contemple el tema *resolución numérica y gráfica de integrales* (Ascheri *et al.*, 2014 y 2015). Dicho software se encuentra disponible en el sitio <http://secanu.exactas.unlpam.edu.ar/>. Se comenzará ahora a trabajar en el desarrollo del software referido al tema *resolución numérica y gráfica de ecuaciones diferenciales ordinarias*, incluido en el curso de Cálculo Numérico que se dicta para las carreras de Profesorado en Matemática (3° Año), Licenciatura en Física (3° Año) e Ingeniería Civil (2° Año) de nuestra Facultad.

Según Alemán de Sánchez (1999), es importante que el software contemple no solamente las prácticas, sino que proporcione al estudiante ayuda en la solución de los problemas y brinde una retroinformación completa, sin limitarse a indicar que se ha cometido un error, sino brindando información acerca del tipo de error. Este y otros aspectos serán considerados e implementados en los diversos ejemplos que se desarrollaran en el software.

Líneas de Investigación, Desarrollo e Innovación

Las líneas de investigación que se abordan con el presente proyecto son las siguientes:

- Enseñanza de matemática, visualización gráfica de diferentes métodos numéricos.
- Inclusión de tecnologías en el proceso de enseñanza y de aprendizaje

Resultados y Objetivos

Ya se ha desarrollado y se ha puesto a punto el software educativo correspondiente a la

resolución numérica y gráfica de integrales. Se propone ahora el logro del siguiente objetivo:

- ✓ Desarrollar el software educativo diseñando las etapas necesarias para lograr el producto, para la *resolución numérica y gráfica de ecuaciones diferenciales ordinarias*.

Para ello, se han planificado las siguientes tareas:

1. Hacer una búsqueda y análisis preliminar de elementos existentes en el campo de la Ingeniería de Software y en el de las Ciencias de la Educación.
2. Utilizar estos elementos, incorporándoles aspectos educativos relativos a la temática involucrada.
3. Experimentar y hacer un análisis y validación de este software.
4. Actualizar, de acuerdo a los resultados obtenidos, el sitio Web de la asignatura "Cálculo Numérico" dependiente de la Facultad de Ciencias Exactas y Naturales de la UNLPam.

Para obtener los datos que permitirán realizar un análisis y validación de esta segunda fase del software educativo elaborado y obtener conclusiones, se utilizarán principalmente la observación, las encuestas y los resultados obtenidos por los estudiantes en las diferentes evaluaciones parciales y finales. Estos resultados constituirán el aporte de una metodología de investigación cuantitativa.

Se desarrollará una estrategia de triangulación que permitirá la coexistencia de la investigación cuantitativa y de la cualitativa (Erickson, 1999). Esta estrategia definida como la combinación de metodologías para el estudio de un mismo fenómeno, considera a los métodos cuantitativos y cualitativos como campos complementarios. Por medio de esta combinación de metodologías, se espera balancear las debilidades de cada uno de los métodos numéricos utilizados y poder fortalecer sus ventajas. También, se espera

encontrar, con la misma metodología, posibles falencias o detectar contradicciones que no se puedan identificar por algunas de las herramientas (observación, encuestas, datos y demás) en forma aislada. En muchos casos, son necesarios ambos tipos de datos (cualitativos y cuantitativos), para una mutua verificación y de forma suplementaria (Vasilachis de Gialdino, 2006).

Un software educativo debe poner énfasis en lograr aprendizajes significativos, brindando posibilidades de vincular los nuevos conceptos con los que se tienen adquiridos y estableciendo relaciones no arbitrarias entre ellos. Debe tener en cuenta los objetivos de las unidades, además de los contenidos y sus relaciones. Debe aportar a la comprensión de las temáticas involucradas. Debe lograr un incremento de la motivación y facilitar el desarrollo de las destrezas. También, debe integrar actividades de evaluación, tanto para el estudiante como para el docente.

Para la elaboración de esta parte del software educativo se utilizarán distintas técnicas, tales como el uso de gráficas por computadoras y animación interactiva para ilustrar y presentar procesos y algoritmos (Alemán de Sánchez, 1999; Cataldi, 1999; Di Battista et al, 1999; Galvis Panqueva, 1992; Marquès, 1996; Price et al, 1998; Rivera Porto, 1997, entre otros).

La visualización de la conducta dinámica de los algoritmos presenta importantes beneficios educativos:

- Logran un incremento de la motivación.
- Facilitan el desarrollo de destrezas.
- Asisten en el desarrollo de habilidades analíticas.
- Ofrecen un buen soporte al docente.
- Permiten la exploración de las peculiaridades de un algoritmo, explorando de forma interactiva.

Las herramientas informáticas y las técnicas numéricas que usaremos son, respectivamente, la computadora y los

métodos numéricos clásicos de resolución de ecuaciones diferenciales ordinarias (Ascheri y Pizarro, 2007; Chapra y Canale, 2007; Gerald y Wheatley, 2000; Mathews y Fink, 2000; Nakamura, 1997), tales como los métodos de Taylor, Euler, Heun, Runge-Kutta, pasos múltiples, entre otros.

A modo de ejemplo, se presenta una de las actividades que se trabajará y resolverá numérica y gráficamente, con el grupo de estudiantes que cursan Cálculo Numérico (de Ingeniería en Construcciones y de Prof. en Matemática), utilizando el software educativo.

Actividad. Una pieza metálica con una masa de 0.1 kg y 200° C (ó 473° K) se coloca en cierto momento dentro de un cuarto con una temperatura de 25° C, en donde está sujeta al enfriamiento por convección natural y la transferencia de calor por radiación. Bajo la hipótesis que la distribución de temperatura es uniforme en el metal, la ecuación de la temperatura se puede escribir como sigue

$$\frac{dT}{dt} = \frac{A}{pcv} [\varepsilon\sigma(294^4 - T^4) + h_c(297 - T)],$$

$$T(0) = 473$$

donde T es la temperatura en grados Kelvin y las constantes son

$p = 300 \text{ kg/m}^3$ densidad del metal; $v = 0.001 \text{ m}^3$ volumen del metal; $A = 0.25 \text{ m}^2$ área de la superficie del metal; $c = 900 \text{ J/kgK}$ calor específico del metal; $\varepsilon = 0.8$ emisividad; $\sigma = 5.67 \times 10^{-8} \text{ w/m}^2\text{K}^4$ cte. Stefan-Boltzmann; $h_c = 30 \text{ w/m}^2\text{k}$ coeficiente de transferencia del calor.

- a) Calcule la temperatura T utilizando el método de Runge-Kutta de cuarto orden para $0 \leq t \leq 180$ (seg) y $h = 1$.
- b) Realice lo mismo que en a) pero utilizando los métodos de Euler y Heun.
- c) Haga los gráficos correspondientes a los distintos métodos de acuerdo a los resultados obtenidos y escriba sus conclusiones.

Formación de Recursos Humanos

Se espera continuar con un grupo de trabajo considerando:

- La motivación de los integrantes a participar de jornadas y eventos científicos.
- La contribución al desarrollo de capacidades y habilidades intelectuales en jóvenes estudiantes que participen o se relacionen con el Proyecto.
- La posibilidad de interesar a otros docentes y estudiantes en participar en actividades de investigación.

A partir de la presentación de los resultados de proyectos relacionados con el actual se han adquirido compromisos tendientes a realizar intercambios de las producciones que surjan, logrando así ampliar la población a la que estará dirigida la implementación del software educativo y que no se restrinja sólo a la cátedra de Cálculo Numérico.

Referencias

Alemán de Sánchez, A. *La enseñanza de la matemática asistida por computadora.* (1999). En línea: <http://www.utp.ac.pa/articulos/enseñarmatematica.html>

Ascheri, M. E.; Pizarro, R. *Libro de Texto para Estudiantes Universitarios: CÁLCULO NUMÉRICO.* EdUNLPam. (2007).

Ascheri, M. E.; Pizarro, R. A.; Astudillo, G.; García, P.; Culla, M. E. *Avances en la Elaboración del Software Educativo para la Resolución Numérica y Gráfica de Integrales,* Memorias de la V REPEM, Vol. 5, ISSN N° 2362-5716, pp. 322-330. (2014).

Ascheri, M. E.; Pizarro, R. A.; Astudillo, G.; García, P.; Culla, M. E.; Pauletti, C., *Utilización de un software educativo para la resolución numérica y gráfica de integrales,* Memorias del XVII Workshop de Investigadores en Ciencias de la

Computación, ISBN N° 978-987-633-134-0, pp. 1-5. (2015).

Cataldi, Z.; Lage, F.; Pessacq, R.; García Martínez, R. *Revisión de Marcos Teóricos Educativos para el Diseño y Uso de Programas Didácticos.* (1999). En línea:

www.itba.edu.ar/capis/webcapis/RGMITBA/comunicacionesrgm/c-icie99-revisione%20marcosteoriciseducativos.pdf

Chapra, S.; Canale, R. *Métodos Numéricos para Ingenieros.* Mc Graw Hill/Interamericana de España, S. A. U. (2007).

Di Battista, P.; Eades, G.; Tamassia, R.; Tollis, I., *Graph Drawing: Algorithms for the Visualization of Graphs.* Prentice Hall. (1999).

Erickson, F. *Métodos cualitativos de investigación sobre la enseñanza.* Paidós. (1999).

Galvis Panqueva, A. *Ingeniería de Software Educativo.* Ediciones Unidades (1992).

Gerald, C. y Wheatley, P. *Análisis Numérico con aplicaciones.* México: Pearson Educación. (2000).

Marquès, P. *El software educativo.* Universidad Autónoma de Barcelona. (1996). En línea: http://www.lmi.ub.es/te/any96/marques_software

Mathews, J.; Fink, K. *Métodos Numéricos con MATLAB.* Prentice Hall. (2000).

Nakamura, S. *Análisis Numérico y Visualización Gráfica con MATLAB.* México: Pearson Educación. (1997).

Price, B.; Beacker, R.; Small, I. *An Introduction to Software Visualization, Software Visualization.* MIT Press. (1998).

Rivera Porto, E., *Aprendizaje Asistido por Computadora. Diseño y Realización.* (1997). En línea:

<http://www.geocities.com/eriverap/libros/Aprend-comp/apen1.html>

Vasilachis de Gialdino, I. *Estrategias de Investigación Cualitativa.* Gedisa, S. A. (2006).

Realidad Aumentada: innovación educativa en las aulas jujeñas

María del Pilar Gálvez^{1,2}, Rudix Camacho^{1,2}, Nélide R. Cáceres¹, Héctor P. Liberatori^{1,2},
Eduardo Graneros², José Quispe¹, Carolina Tolaba¹, Evelina C. Velázquez,
Daniel A. Lamas¹, Brenda N. Veramendi¹

¹Facultad de Ingeniería, Universidad Nacional de Jujuy
Italo Palanca 20 San Salvador de Jujuy – 0388 4221576

²Universidad Católica de Santiago del Estero – Departamento Académico San Salvador
Lavalle 333 San Salvador de Jujuy – 0388 4236139
mdpgalvezdiaz@fi.unju.edu.ar

Resumen

Las computadoras y las nuevas tecnologías en la educación pueden ser vistas como herramientas y materiales de construcción que facilitan el aprendizaje y el desarrollo de habilidades; constituyendo distintas formas de aprender de los alumnos y no deben ser consideradas como un fin, sino como un medio. Se busca utilizar estas herramientas para promover, facilitar y enriquecer la comprensión de temas y conceptos propios e importantes de las diferentes asignaturas curriculares dentro del proceso educativo.

Este trabajo se centra en la introducción y aplicación de Realidad Aumentada como nueva tecnología que puede incorporarse en el proceso de enseñanza y aprendizaje dentro del sistema Educativo, mediante la combinación de diferentes tipos de materiales (auditivos, visuales y audiovisuales).

Palabras clave: TIC's – Realidad Aumentada – Educación.

Contexto

La presente investigación se enmarca en el proyecto “Realidad Aumentada: innovación educativa en las aulas

jujeñas”, cuyo objetivo principal es introducir los aportes de la tecnología de la Realidad Aumentada en las prácticas educativas de forma que permita favorecer aprendizajes significativos en los estudiantes.

Este proyecto, categoría A, está aprobado por la Secretaría de Ciencia y Técnica de la Universidad Nacional de Jujuy y se encuentra bajo incentivo.

Participa en la coordinación de este proyecto la Universidad Católica de Santiago del Estero, Departamento Académico San Salvador. Con la incorporación de docentes investigadores de la mencionada casa de estudios, se constituyó un equipo de trabajo más amplio, diversificado, interdisciplinario beneficiando al proyecto con el aporte del área de conocimiento de todos sus integrantes.

Introducción

Los sistemas educativos, desarrollados por más de doscientos años, se enfrentan hoy la necesidad de cambios sustantivos. Las nuevas condiciones sociales, culturales y tecnológicas tensionan la necesidad de realizar cambios urgentes en las instituciones educativas y sus prácticas pero las nuevas prácticas educativas no serán posibles en los ambientes escolares tradicionales. En este

sentido, actualmente se percibe: la fragmentación del sistema educativo, la pérdida del sentido de la tarea pedagógica, el desencuentro entre los modelos educativos familiares y la cultura escolar, la erosión de la autoridad docente, la proliferación de miradas estigmatizadoras sobre los jóvenes, etc. Estos desdibujamientos trazan nuevos contornos de acción para la práctica educativa y suponen un reto para quienes intentan asumir la transmisión educativa de los contenidos de la cultura. Coincidimos con Larry Cuban (2011) en reconocer que "sin poner atención a las condiciones del lugar de trabajo de los docentes y sin el reconocimiento de los saberes que traen los alumnos al aula, hay poca esperanza de que las nuevas tecnologías tengan más que un mínimo impacto en la enseñanza y el aprendizaje".

Las instituciones educativas y sus prácticas han quedado atrapadas en el tiempo, en el sentido de su oferta cultural, de aprendizaje y de enseñanza frente a nuevos sujetos producto de las condiciones sociales, culturales y fundamentalmente tecnológicas, como son los estudiantes que corresponden a la denominada generación NETs (Córlica, 2009). Este autor indica, que los alumnos NETs, al estar expuestos a mucha información en las redes cualquier situación de enseñanza puede resultarles aburrida y continuamente buscan ámbitos y experiencias de aprendizaje divertidas. Requieren de instrucciones precisas, de planes detallados de trabajo que impliquen recompensas inmediatas, en consecuencia la responsabilidad y compromiso surgen cuando encuentran sentido en lo que hacen, cuando son útiles para alguien o para algo. Frente a este escenario, la oferta educativa debería abandonar su marca tradicional y apuntar a la generación de saberes y formas de

enseñarlos que tomen en cuenta las características de estos nuevos sujetos que habitan las aulas

Hemos entrado en un período caracterizado por nuevas formas de representación del conocimiento que afectan la manera misma en que se crean y se organizan los contenidos, como también las muchas formas en que se distribuyen los conocimientos. Las aplicaciones de multimedia aún no están presentes en la cotidianidad educacional, pero paulatinamente será parte de la vida diaria en las instituciones educativas. En este sentido las aplicaciones de las Tics no sólo se constituyen como una tecnología novedosa sino también como una nueva forma de representar conocimientos y actividades y dar respuesta a las demandas de aprendizaje de las nuevas generaciones de estudiantes que pueblan las aulas

Los materiales didácticos que operan con el lenguaje multimedial se acercan más a la experiencia de los estudiantes, lo que conlleva un aumento en la capacidad de retención de la información y una mejora en los resultados pedagógicos. Las computadoras y las nuevas tecnologías en la educación pueden ser vistas como herramientas y materiales de construcción que facilitan el aprendizaje y el desarrollo de habilidades; constituyendo distintas formas de aprender de los alumnos y no deben ser consideradas como un fin, sino como un medio. Se busca utilizar estas herramientas para promover, facilitar y enriquecer la comprensión de temas y conceptos propios e importantes de las diferentes asignaturas curriculares dentro del proceso educativo.

Este trabajo se centra en la introducción y aplicación de Realidad Aumentada (RA) como nueva tecnología que puede incorporarse en el proceso de enseñanza y aprendizaje dentro del sistema Educativo, mediante la

combinación de diferentes tipos de materiales (auditivos, visuales y audiovisuales).

La Realidad Aumentada propone superponer o añadir, en tiempo real, elementos virtuales al campo de visión de una persona, de modo tal que la información requerida esté presente de una manera natural para el usuario. Estos elementos consisten en objetos, sonidos, imágenes y textos visuales generados por computadora, llevando a que el usuario no se sumerja completamente en un mundo virtual sino en una mezcla de éste con el mundo real, manteniéndolo en contacto con la realidad y en ocasiones permitiéndole interactuar con objetos virtuales. Esto es diferente de la Realidad Virtual en donde se genera un mundo artificial el que reemplaza por completo al mundo real, teniendo por objetivo principal que el usuario pueda interactuar y sentirse inmerso dentro de éste mundo generado. (Abásolo et al., 2008)

Los sistemas de Realidad Aumentada se diferencian de los sistemas tradicionales en la interacción más intensa con el usuario y con el entorno para la realización de una tarea.

Según Azuma (2001) un sistema de RA tiene las siguientes propiedades: combina objetos reales y virtuales en un entorno real; corre interactivamente en tiempo real y lo virtual está registrado en el mundo real.

Algunos de los sistemas que implementan la RA hacen uso de un "marcador" que consiste generalmente en una imagen impresa. El uso de marcadores es una de las técnicas de interacción utilizadas en la RA. Básicamente estos marcadores se exhiben delante de una cámara y un software procesa la imagen para que muestre información virtual sobre un "display", añadiéndose sobre el escenario real, este proceso se llama registro de imágenes. La

RA se apoya en marcadores para que la computadora tenga un punto de referencia sobre el cual posicionar y orientar el objeto de modo que la información virtual sea alineada con el escenario del mundo real, (Cardoso, 2004).

Para hacer uso de esta tecnología se debe contar mínimamente con: (Fig. 1)

Cámara Web: necesaria para la captura de las escenas del mundo real. Es el dispositivo de captura de video más accesible y de mayor disponibilidad para la implementación de la Tecnología de RA, la mayoría de las computadoras las tienen incorporadas y cuentan con un bajo precio en el mercado. Actúa como puente entre el ambiente real y el software de RA. Debe contar con libertad de movimiento, buenas características de resolución y un ambiente iluminado para detectar con facilidad el patrón impreso.

Software de Realidad Aumentada: la cámara web genera un video del entorno existente en tiempo real enviándolo al software de RA, quien es el encargado de procesar la información adquirida por la cámara, los datos almacenados en su base de datos, la información proporcionada por los marcadores así como la posición, orientación y marcas de reconocimiento. Una vez procesada toda esta información se procede a una adecuada fusión entre imágenes reales y objetos virtuales. Finalmente, sobre la imagen capturada y mediante librerías externas, serán dibujados los objetos 3D de modo que aparezcan sobre el patrón en la posición, orientación y tamaño correspondiente al punto de vista de la cámara, (Kato et al., 1999.)

Dispositivo de Visualización: a través de éste, pantalla o monitor, se proyectará la suma de lo real y lo virtual, lo cual consiste en añadir información de forma numérica, textual o bien objetos que no estaban presentes en la escena original, conformando la RA. Esta composición es

forma se podrá evaluar si la experiencia cumple con los objetivos propuestos.

Objetivos

Este proyecto tiene estipulados cuatro años de duración (2016-2019) y los siguientes objetivos principales:

-Introducir los aportes de la tecnología de la Realidad Aumentada en las prácticas educativas con el propósito de favorecer aprendizajes significativos en los estudiantes.

-Propiciar el desarrollo de aplicaciones de RA como innovaciones educativas y como recursos metodológicos novedosos para los procesos de enseñanza y aprendizaje.

- Promover instancias de capacitación e intercambio entre docentes e interesados que puedan requerir de las herramientas de RA desarrolladas por el grupo de investigación.

Estructura del equipo de investigación

Directora: Mg. María del Pilar Gálvez.
Categoría de Investigación III.

Codirectora: Mg. Rudix, Camacho.
Categoría de Investigación III.

Investigadores:

- Ing. Nélide R. Cáceres. Categoría de Investigación V.
- Mg. Héctor Liberatori. Categoría de Investigación IV.
- Ing. Eduardo Graneros.
- Ing. Ana Carolina Tolaba.
- Ing. José Quispe.

Profesionales: Ing. Informáticos:

- Evelina Carola Velázquez
Alumnos colaboradores de la carrera Ingeniería Informática:
- Daniel Alberto Lamas.
- Brenda Noelia Veramendi.

Referencias

Abásolo, María J., Rocío A, Rodríguez. *Aplicaciones de la Realidad Aumentada: Enfoques Posibles para Entornos Educativos*. La Plata, 2008.

Azuma, R., Yohan, B., Reinhold, B., Steven, F., Simon, J. & M. Blair. *Recent Advances in Augmented Reality*. IEEE Computer Graphics and Applications. 2001.

Billinghurst, M & H. Kato, and I. Poupyrev. *The MagicBook: a transitional AR interface*. Computers & Graphics, vol. 25, núm. 5, pp. 745-753. 2001.

Cardoso, A. & Jr. Lamounier. *Artoolkit Aspectos Técnicos e Aplicações Educacionais*. San Pablo, 2004.

Carracedo, J. de Pedro, C. L. Martínez Méndez. *Realidad Aumentada: Una Alternativa Metodológica en la Educación Primaria Nicaragüense*. IEEE-RITA Vol. 7, Núm. 2. 2012.

Córica, José L., Patricia Dinerstein. *Diseño Curricular y Nuevas Generaciones: Incorporando a la Generación Net*. ISBN 978-987 -24871 -2-6. Editorial Virtual Argentina. Mendoza, 2009.

Cuban, Larry. The Myth of Failed School Reform. Disponible en: <http://larrycuban.wordpress.com/2011/02/02/the-myth-of-failed-school-reform-part-1/> y <http://larrycuban.wordpress.com/2011/02/04/myth-of-failed-school-reform-part-2/>

Kato, H., & M. Billinghurst. *Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System*. 2do Workshop Internacional sobre Realidad Aumentada. San Francisco, 1999.

Propuesta de un Marco Conceptual para el Diseño e Implementación de Repositorios Institucionales de Contexto Educativo Ubicuos (RICEU)

Lic. Guillermo Javier Lafuente¹, Mg. José Luis Filippi², Lic. Gustavo Lafuente³, Mg. Carlos Ballesteros⁴

GIAU⁵ – Facultad de Ingeniería – UNLPam.

Calle 110 esq. 9 n° 390

{lafuente¹, filippij², gustavo³,balleste⁴}@ing.unlpam.edu.ar

⁵Grupo de Investigación de Ambientes Ubicuos – <http://giau.ing.unlpam.edu.ar/>

Resumen

Este trabajo tiene como objetivo definir estrategias para el diseño e implementación de Repositorios Institucionales (RIs) de características ubicuas. Nos centramos principalmente en el contexto de uso, y desde ese lugar proyectar soluciones que se adapten al entorno de quien consulta un repositorio. Tenemos como meta bajo esta línea, lograr definir un marco conceptual que sirva de guía a los desarrolladores para el proceso de desarrollo de Repositorios Institucionales de Contexto Educativo Ubicuos (RICEU) basados en la Web Ubicua [1].

Palabras clave: Repositorios Institucionales de Contexto Educativo, Web Ubicua, Interfaces adaptativas.

Contexto

La actual línea de trabajo se lleva a cabo en el marco del proyecto de I+D denominado “*Tecnología Móvil Aplicada en la Educación*”, proyecto aprobado por resolución N° 122/14 del Consejo Directivo de la Facultad de Ingeniería de la UNLPam, dentro del grupo GIAU⁵.

Introducción

Esta línea de I+D intenta definir las guías básicas para la creación, organización y desarrollo de Repositorios Institucionales de Contexto Educativo Ubicuos (RICEU) de Acceso Abierto. En ese sentido, es de interés poder definir un marco de trabajo “framework” que sirva a los ingenieros de software como estrategia conceptual para el desarrollo de RI de contenido educativo, principalmente orientado a contextos universitarios o de características similares. Se pretende además, que dicho marco pueda ser la base para construir RI con características ubicuas. Se intenta lograr a través de la implementación del marco, que los RI se desarrollen y adapten al contexto de uso de quien los consulta de manera autónoma a través de la Web.

Las Universidades se ven influenciadas por un contexto que cambia de forma permanente, lo que afecta de manera sustancial sus funciones principales: investigación, enseñanza y la transmisión del conocimiento, entre otras. En cuanto a su dimensión educativa, vemos una inclinación hacia metodologías didácticas más constructivistas, centradas en “aprender a aprender” y en el aprendizaje autónomo y

para toda la vida, y los cambios asociados en los roles de docentes y alumnos, con respecto a los espacios y servicios [2] van cambiando todo el tiempo. Si hay que elegir un aspecto clave de todas estas transformaciones, el motor de cambio sin lugar a duda ha sido la influencia de las TICs, y como éstas han logrado transformar la forma en la que se genera, se obtiene, y se utiliza la información y el conocimiento necesario para llevar a cabo estas actividades.

Las nuevas propuestas educativas a distancia han generado nuevas tendencias tecnológicas en las bibliotecas universitarias, por un lado el desarrollo de nuevos servicios mediados por TICs, por otro, el acceso a las colecciones digitales además de la aplicación de tecnologías emergentes en el acceso y recuperación de la información. Estos cambios han generado que la biblioteca ubicua [3] sea una realidad y brinde servicios en cualquier momento y en cualquier lugar.

Para implementar un RI ubicuo es preciso que entendamos los aspectos de la computación ubicua y focalizarnos más precisamente en las aplicaciones Web Ubicuas. Pues los RI actuales no dejan de ser aplicaciones Web accedidas a través de la red.

Una Aplicación Web Ubicua [4] es una aplicación Web que sufre el síndrome: *en cualquier momento / en cualquier lugar / en cualquier medio*. Esto significa que una aplicación web ubicua debería estar diseñada desde el principio teniendo en cuenta no sólo su naturaleza hipermedia, sino también el hecho de que debe ejecutarse "tal cual" en una variedad de plataformas, incluyendo dispositivos móviles, computadoras de escritorio, Notebook, televisores, etc. Esto implica que una aplicación web basada en el concepto de "en todas partes", debe tener en cuenta las diferentes capacidades de los dispositivos

que comprenden el tamaño de la pantalla, el tamaño de almacenamiento local, el método de entrada, la capacidad de la red, etc. Se ofrecen nuevas oportunidades en términos basados en la geolocalización, en el tiempo, y en los servicios personalizados, teniendo en cuenta las necesidades y preferencias de los usuarios particulares. En consecuencia, una aplicación de web ubicua debe ser, por un lado, sensible al contexto, es decir, sensible al medio ambiente que se está ejecutando, y por el otro lado, debe ser compatible para soportar la personalización.

Actualmente la mayoría de los RI son sistemas basados en la Web orientados principalmente para ser accedidos a través de un navegador Web e ideados para acceder desde una notebook o PC. En tal sentido, proponer una estrategia conceptual que ayude a quienes implementen RIs ubicuos constituye un aporte significativo en la materia, pues se trata de brindar un marco conceptual que permita a los diseñadores disponer de una herramienta para desarrollar RIs adaptables a distintas interfaces conforme se trate del entorno que rodea al usuario.

Como se indicó anteriormente, este trabajo se centra en RIs basados en la web y desde el punto de vista de la ubicuidad, nos focalizamos en la adaptabilidad automática de la interfaz de usuario conforme al contexto de uso. Tal supone la W3C [1], mucha es la complejidad de las aplicaciones web ubicuas. Abordaremos la ubicuidad desde el punto de vista de la adaptación automática de la Interface de Usuario (IU) y consideraremos de interés, las interfaces de usuarios basadas en modelos (Model-based UI) [5].

Los desarrolladores de interfaz de usuario para sistemas interactivos tienen que hacer frente a múltiples fuentes de heterogeneidad:

- De los usuarios finales. Éstos difieren con respecto a sus preferencias, capacidades, cultura y el nivel de experiencia.
- De las plataformas informáticas, las modalidades de interacción y las capacidades de entrada / Salida.
- De los lenguajes de programación / de marcado y el conjunto de herramientas (widgets o librerías).
- De los entornos de trabajo. Los desarrolladores tienen que tener en cuenta diferentes limitaciones contextuales (por ejemplo, entornos ruidosos, movilidad, etc.).
- La variabilidad del contexto de uso. Además de ser heterogéneo, el contexto del uso evoluciona dinámicamente, llamando así a las interfaces de usuario “plásticas”, es decir, interfaces de usuario capaces de adaptarse, preservando los valores humanos [6].

El desarrollo del modelo de interfaz basada en usuario (MBUID) constituye un enfoque que tiene como objetivo hacer frente a los desafíos mencionados al disminuir el esfuerzo necesario para desarrollar interfaces de usuario al tiempo que garantiza su calidad. El propósito del diseño basado en modelos es identificar los modelos de alto nivel que permiten a los diseñadores especificar y analizar las aplicaciones de software interactivas en un nivel más orientado a la semántica, en lugar de comenzar de inmediato para hacer frente al nivel de aplicación.

Por consiguiente, el desafío es avanzar en un marco de trabajo para desarrollar RI basados en interfaces de usuario multi-contexto garantizando la ubicuidad.

Líneas de Investigación

Para construir un RI se necesita definir una arquitectura en capas, en la que estén presentes todos los aspectos

arquitectónicos para desarrollarlo. Partiendo de una estructura cliente/servidor basada en la Web se diseñará la estrategia conceptual definiendo los aspectos a tener en cuenta para las interfaces ubicuas, como así también los aspectos propios del dominio arquitectónico de los RIs.

En particular, será relevante el análisis, adaptación y uso de modelos y arquitecturas para la creación de repositorios. Se avanza en el estudio de líneas de trabajo basadas en los modelos propuestos por Bawden y Rowlands [7], y el modelo de referencia OAIS (Norma ISO:14721) [8]. También son considerados los modelos “*Streams, structures, spaces, scenarios, societies (5S)*” [9], el modelo DELOS [10], el modelo conceptual FRBR (*Functional Requirements for Bibliographic Records*) [11], y el metamodelo CRADLE (*Cooperative-Relational Approach to Digital Library Environments*) [12], entre otros. Por otro lado es de destacar la propuesta de De La Fuente y Pérez [13] presentando un modelo de Repositorio Institucional de Contenido Educativo (RICE). La propuesta apunta hacia un modelo de repositorio institucional que permita configurar y administrar la colección de activos digitales, sustentado en la integración e interoperabilidad de contenidos y sistemas digitales educativos de una institución y que pueda servir de ejemplo a las bibliotecas universitarias que se propongan afrontar y atender esta necesidad.

No obstante, y sobre esta línea de trabajo en particular, nos ubicaremos en las interfaces adaptativas, y estudiaremos distintos marcos de referencia desarrollados para capturar conceptualmente los aspectos importantes de un proceso de MBUID. Szekely introduce una arquitectura genérica para MBUID [14]. Da Silva describe una

arquitectura para el desarrollo de la interfaz de usuario utilizando un enfoque MBUID [15]. Calvary [16] propone la primera versión de un framework para múltiples contextos de uso de interfaces de usuario con un enfoque basado en modelos, que luego amplió con las relaciones y definiciones adicionales para dar lugar a un marco de referencia [17, 18]. Por otro lado, Collignon y Coutaz [19, 20] plantean un enfoque de ingeniería de software basada en modelos de interfaces de usuario plástica multi-propósito.

Resultados y Objetivos

El objetivo general de esta línea de investigación es definir estrategias para el diseño e implementación de Repositorios Institucionales de Contexto Educativo Ubicuos (RICEU). Centrándonos en el contexto de uso y desde ese lugar proyectar soluciones que se adapten al entorno de quien lo utiliza. Para ello, se pretende desarrollar un marco conceptual de trabajo que sirva a los ingenieros de software como guía para la puesta en marcha de RICEUs. El aspecto de ubicuidad es analizado desde el punto de vista de la adaptabilidad de la Interfaz de Usuario (IU) y bajo una arquitectura que siga las premisas propuestas en la W3C respecto a la Web Ubicua.

Por un lado se está trabajando con el modelo de referencia OAIS [8], para extenderlo al contexto de los RICEUs teniendo en cuenta la propuesta del modelo RICE [13]. Por otro lado, hemos comenzado con el estudio de marcos de trabajo para el desarrollo de interfaces adaptativas como el propuesto por Clavary [15], Collignon et. all. [18] y Coutaz et. all. [19] a fin de adaptarlo al dominio de RIs, y que su estrategia de implementación de IU adaptativa forme

parte de un marco conceptual más integral.

Pensamos que este aporte puede contribuir sustancialmente para el desarrollo de RICEUs, pudiendo adaptarse a distintos niveles educativos, no sólo el universitario. Para ello, es de suma importancia lograr un marco conceptual bien definido para conducir a los actores del desarrollo de software en la puesta a punto de este tipo de repositorios basados en la Web y que a partir de su uso se pueda lograr una sola implementación adaptable a distintos ámbitos y contextos educativos y bajo distintas condiciones según se lo requiera.

Formación de Recursos Humanos

Actualmente, el proyecto cuenta con la siguiente conformación de grupo de trabajo: un Director de Proyecto y un Co-Director, tres Investigadores, un Asistente de Investigador y dos Tesistas. Bajo la línea de investigación presentada en este artículo, se está desarrollando una tesis de posgrado y una tesis de grado. En la de postgrado se pretende avanzar en la línea de trabajo actual. En cuanto a las tesis de grado, tienen como objetivo avanzar en la implementación de un repositorio institucional de acceso abierto para la Facultad de Ingeniería de la UNLPam de características ubicuas soportado a través del marco conceptual propuesto.

Referencias

- [1] W3C, (2014), “*Ubiquitous Web Applications Activity Statement*”, <http://www.w3.org/2007/uwa/Activity.html>
- [2]. Bueno de la Fuente, Gema; Hernández Pérez, Tony (2011). “*Estrategias para el éxito de los repositorios institucionales de contenido educativo en las bibliotecas digitales universitarias*”. BiD: textos universitaris de

biblioteconomia i documentació, Junio, núm. 26.
<http://www.ub.edu/bid/26/bueno2.htm>

[3] S. G. Martín, “*La biblioteca ubicua: servicios en cualquier momento y en cualquier lugar.*” PROED. (Programa de Educación a Distancia). Secretaría de Asuntos Académicos, Universidad Nacional de Córdoba, 15-Mar-2013.

[4] Finkelstein, A. et al; (2002). “*Ubiquitous Web Application Development: a framework for understanding*”, 6th World Multiconference on Systematics, Cybernetics and Informatics, <http://www.cs.ucl.ac.uk/staff/A.Finkelstein/papers/uwa.pdf>

[5] W3C Working Group, (2014), “*Introduction to Model-Based User Interfaces*”, <http://www.w3.org/TR/2014/NOTE-mbui-intro-20140107/>

[6] Thevenin, D., Coutaz, J.; (1999), “*Plasticity of User Interfaces: Framework and Research Agenda*”. In Proc. Interact99, Edinburgh, A. Sasse & C. Johnson Eds, IFIP IOS Press Publ., pp.110-117.

[7] D. Bawden and I. Rowlands, “*Digital libraries: assumptions and concepts*,” Libri, 1999. <http://discovery.ucl.ac.uk/166226/>

[8] CCSDS, (2002), “*Reference Model for an Open Archival Information System (OAIS)*”, ISO 14721.

[9] M. A. Gonçalves, E. A. Fox, L. T. Watson, and N. A. Kipp, (2004), “*Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries*” ACM Trans. Inf. Syst., vol. 22, no. 2, pp. 270–312, Apr. 2004.

[10] Leonardo Candela, Castelli, et. al., (2007), “*Setting the Foundations of Digital Libraries*,” D-Lib, vol. 13, no. 3/4, Mar. 2007.

[11] FRBR, (2009), “*Functional Requirements for Bibliographic Records - IFLA.*”

[12] A. Malizia, P. Bottoni, and S. Levialdi, (2010), “*Generating Collaborative Systems for Digital Libraries: a Model-Driven Approach*”

Information Technology & Libraries, <http://ec2-50-19-240-191.compute-1.amazonaws.com/1330/>.

[13] G. de la Fuente and D. Pérez, “Modelo de repositorio institucional de contenido educativo (RICE): la gestión de materiales digitales de docencia y aprendizaje en la biblioteca universitaria,” Univ. Carlos III Madrid. ..., 2010.

[14] Szekeley, P., (1996), “*Retrospective and Challenges for Model-Based Interface Development*”. Proc. of the 3rd International Eurographics Workshop, pp. 1-27.

[15] Da Silva, P. P., (2000), “*User Interface Declarative Models and Development Environments: A Survey*”. Proc. of the 7th International Conference on Design, Specification, and Verification of Interactive Systems, pp. 207-226.

[16] Calvary, G., Coutaz, J., Thevenin, D., (2001), “*A Unifying Reference Framework for the Development of Plastic User Interfaces*”. Proc. of EHCI, pp. 173-192.

[17] Calvary, G. et al., (2002), “*The CAMELEON Reference Framework*”, CAMELEON Project, <http://giove.isti.cnr.it/projects/cameleon/pdf/CAMELEON%20D1.1RefFramework.pdf>.

[18] Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonck, J., A , (2003), “*Unifying Reference Framework for Multi-Target User Interfaces, Interacting with Computers*”, Vol. 15, No. 3, pp. 289-308.

[19] B. Collignon, J. Vanderdonck, and G. Calvary, “*Model-driven engineering of multi-target plastic user interfaces*,” in Proceedings - 4th International Conference on Autonomic and Autonomous Systems, ICAS 2008, 2008, pp. 7–14.

[20] J. Coutaz and G. Calvary, “HCI and Software Engineering for User Interface Plasticity,” Human-Computer Interact. Handb. Fundam. Evol. Technol. Emerg. Appl. Third Ed., p. 1195-1220, 2012.

M-library: Servicios de una biblioteca universitaria disponibles a través de dispositivos móviles

Lic. Guillermo Javier Lafuente¹, Mg. José Luis Filippi², Ap. Enzo D. Sanchez³
GIAU⁴ – Facultad de Ingeniería – UNLPam.

Calle 110 esq. 9 n° 390

{lafuente¹, filippij²}@ing.unlpam.edu.ar

enzo214@hotmail.com³

⁴Grupo de Investigación de Ambientes Ubicuos – <http://giau.ing.unlpam.edu.ar/>

Resumen

Bajo el concepto de M-library, se pretende extender los servicios de una biblioteca universitaria haciéndolos disponibles a través de dispositivos móviles para que los integrantes de una comunidad educativa puedan accederlos desde cualquier lugar y en cualquier momento. El desarrollo de un aplicativo de estas características estará ligado a una serie de otros aplicativos que la Facultad de Ingeniería de la UNLPam se encuentra desarrollando bajo el proyecto de I+D denominado *Tecnología Móvil Aplicada a la Educación*. Se pretende integrarlo a una arquitectura robusta basada en el modelo cliente/servidor que sirva a los objetivos del aprendizaje móvil (m-learning) que la institución mencionada intenta implementar.

Palabras clave: M-library, Librería Móvil, Web Ubicua, Web Móvil, Interfaces Adaptativas.

Contexto

La actual línea de trabajo se lleva a cabo en el marco del proyecto de I+D denominado *“Tecnología Móvil Aplicada*

en la Educación”, proyecto aprobado por Resolución N° 122/14 del Consejo Directivo de la Facultad de Ingeniería de la UNLPam, dentro del grupo GIAU (Grupo de Investigación de Ambientes Ubicuos).

Introducción

La aparición continua de dispositivos electrónicos cada vez más pequeños e inteligentes, invita a reflexionar sobre una nueva forma de hacer la educación convirtiendo el aula tradicional en ambientes inteligentes, en el cual los actores involucrados puedan recibir variados servicios de educación, conocido como Mobile Learning (M-learning).

Se puede definir el M-learning como una nueva forma de educación creada a partir de la conjunción entre el e-learning y la utilización de los dispositivos móviles inteligentes, y que se fundamenta en la posibilidad que ofrecen estos nuevos dispositivos, de combinar la movilidad geográfica con la virtual, lo cual permite aprender dentro de un contexto, en el momento en que se necesita y explorando y solicitando la información precisa que se requiera saber [1]. Un ambiente inteligente es un espacio físico en el cual todo está conectado con todo, donde la

información está presente en forma instantánea, en el cual las computadoras y las personas se comunican sin fronteras. Bajo esta modalidad se intenta perseguir el desarrollo de habilidades cognitivas por parte de los educandos [2].

La Web Móvil se define como una red en la que los usuarios pueden acceder a la información basada en la Web, desde cualquier lugar, sin importar el tipo de dispositivo utilizado [3]. El uso de dispositivos móviles para acceder a Internet y la búsqueda de información están creciendo considerablemente. Las mejoras tecnológicas, tales como teléfonos móviles, la conectividad a bajo costo y la transmisión de datos a alta velocidad se encuentran entre los factores más importantes que han llevado a la creciente utilización de dispositivos móviles [3] para distintas tareas.

El acceso a Internet desde dispositivos móviles ha cambiado no sólo la forma de comunicarse, sino que también ha afectado a la forma en que los usuarios buscan, recuperan y utilizan la información para hacer frente a sus necesidades diarias. Esto fue reconocido por los bibliotecarios, que vieron a sus usuarios utilizando sus teléfonos móviles en lugar de computadoras portátiles y computadoras de escritorio para buscar en el catálogo, comprobar los horarios de apertura de la biblioteca y mantener contacto con el personal de la misma [4]. Los bibliotecarios rápidamente se dieron cuenta de la necesidad de ofrecer sitios web para brindar servicios de la biblioteca, que podrían ser buscados fácilmente a través de dispositivos móviles.

El concepto de la biblioteca móvil (m-library) no es nuevo. Se refiere en primer lugar a aquellos vehículos diseñados específicamente para su uso como una biblioteca [5] y que transportan material bibliográfico. Este tipo de biblioteca fue

destinada a proporcionar servicios a zonas en donde la biblioteca tradicional no llegaba, como en pequeñas aldeas y barrios que no contaban con un edificio con servicios de biblioteca. Este tipo de biblioteca, también llamada la biblioteca bibliobús, –término que se utiliza indistintamente– funcionaba como una extensión de la biblioteca tradicional. Con el advenimiento de la tecnología móvil, el término de biblioteca móvil fue pronto utilizado para referirse a los servicios de información ofrecidos a través de los dispositivos móviles [6,7,8].

En tal sentido, esta propuesta pretende indagar en los servicios [9,10,11] que se pueden brindar bajo esta modalidad, por parte de una biblioteca a través de la diversa gama de dispositivos móviles que existen en la actualidad en el mercado. En especial se pone énfasis en el estudio de las demandas actuales por parte de los actores principales de una institución universitaria detectando las necesidades a cubrir, al igual que el estudio y disponibilidad de dispositivos móviles por parte de esos actores, pretendiendo lograr un servicio de m-library adecuado a esas demandas [12].

Pensar en una estrategia de implementación enmarcada en el concepto de interfaz adaptativa, centrada en un framework para el desarrollo de interfaces adaptativas [13,14] permitirá optimizar los tiempos de trabajo, orientando el producto resultante a ser ejecutado en diversas pantallas y plataformas. De esta forma se intenta lograr bajo este enfoque de diseño [15], la implementación de una sola vez de los servicios en línea para una biblioteca y que todo su contenido pueda ser accesible siempre desde cualquier lugar y desde cualquier dispositivo o plataforma en el que se opere, tan solo disponiendo una conexión a internet.

Llevar a cabo el servicio de m-library, e integrarlo a una arquitectura basada en la estrategia de m-learning la cual se viene desarrollando por parte del proyecto de I+D mencionado al principio, permitirá ampliar la gama de aplicativos disponibles para el desarrollo de los procesos de enseñanza y aprendizaje, extendiendo aún más los servicios que una institución, en este caso universitaria, pueda ofrecérselos a sus actores principales ya sean docentes, investigadores, graduados y estudiantes entre otros.

Líneas de Investigación

En el marco del proyecto I+D indicado anteriormente, se espera avanzar en el ámbito de interfaces adaptativas basadas en la Web para dispositivos móviles basadas en el contexto del usuario destinadas al ámbito educativo. Se pretende direccionar el trabajo hacia una arquitectura centrada en un modelo cliente/servidor orientada a servicios para usuarios que se desarrollen en el contexto educativo universitario. Conforme a este enfoque se desea desarrollar un nuevo servicio en el marco de una propuesta de arquitectura mayor de servicios, soportada bajo el concepto de m-learning, la cual viene siendo desarrollada en el proyecto de I+D. Este nuevo servicio está orientado al ámbito de las bibliotecas móviles (m-library) constituyéndose en una extensión más de la biblioteca tradicional.

Para avanzar sobre el trabajo se ha realizado un plan de actividades con ciertas líneas de investigación a fin de poder alcanzar los objetivos deseados. A continuación enumeramos algunas de ellas:

a. Indagar en las nuevas tecnologías de desarrollo para aplicaciones Web conforme al nuevo estándar HTML5.

b. Analizar en las nuevas recomendaciones de la W3C para el desarrollo de interfaces para aplicaciones Web móviles, a saber:

b.1 CSS Flexible Box Layout Module Level 1 [16]

b.2 Indexed Database API [17] y Realm Database móvil [18]

b.3 CSS Grid Layout Module Level 1 [19]

c. Estudiar los marcos de trabajo propuestos para el desarrollo de Interfaces Adaptativas centradas en el contexto de uso.

d. Estudiar marcos de trabajo para el desarrollo de servicios basados en m-library y m-learning.

e. Implementar un prototipo de aplicación móvil conforme a las tecnologías indicadas anteriormente en el marco del proyecto de I+D (m-library).

e. Desarrollar en paralelo un informe de tesis final de carrera en función del tema indicado para este plan de actividades.

e.1 Publicación del avance en congresos y/o revistas científicas.

Durante el avance de estas actividades, se documentará el trabajo a través del desarrollo de una tesis (la cual servirá como proyecto final de carrera de uno de los becarios) como así también el desarrollo de trabajos científicos que se originen del producto resultante.

Resultados y Objetivos

El objetivo de esta línea de trabajo es avanzar en el desarrollo de interfaces adaptativas orientadas a dispositivos móviles centradas en el contexto de uso, analizando distintas tecnologías para llevarlas a cabo. Estas actividades

constituyen un aporte sustancial al proyecto de I+D en pos de poder concretar aplicaciones web móviles, capaces de adaptarse a distintos dispositivos tecnológicos de forma automática. Para ello, se pretende desarrollar un prototipo de un aplicativo orientado a dispositivos móviles, en especial, basado en el dominio de bibliotecas móviles. El desarrollo de este prototipo, centrado en la estrategia de Interfaces Adaptativas [20], contribuirá para la definición y construcción de una arquitectura más robusta orientada en el modelo cliente/servidor denominada m-university (universidad móvil), la cual contemple una serie de servicios destinados a los usuarios de una comunidad académica, disponibles en cualquier dispositivo móvil, independientemente de su tamaño de pantalla y plataforma.

Como resultado de esta línea de trabajo podemos indicar que en la actualidad se está discutiendo el conjunto de requisitos funcionales que llevará soportado el aplicativo móvil. A continuación resaltamos algunos de ellos:

- Servicio de notificaciones de nuevos recursos (bibliográficos y digitales) basados en wishlist o listas de interés.
- Disponibilidad de préstamos y/o descarga de material bibliográfico
- Reservas de recursos bibliográficos para su uso.
- Consultas en el acervo bibliográfico.
- Realidad Aumentada para el recurso digital accedido.

Paralelamente se está discutiendo la plataforma de soporte en la que estará destinado el aplicativo. En tal sentido, se está trabajando en dos tipos de estrategias en la que se soportarán los requerimientos de m-library, una integrada al conjunto de

servicios que brindará un aplicativo más grande basado en el concepto de m-university sobre dispositivos Android, y otra plataforma, basada en la web adaptativa que funcione como servicio independiente, y destinada ejecutarse en cualquier dispositivos móvil. A partir de ello se avanzará en futuras etapas que involucre el proceso de desarrollo del servicio de m-library de acuerdo a la planificación prevista.

Formación de Recursos Humanos

Actualmente, el proyecto cuenta con la siguiente conformación de grupo de trabajo: un Director de Proyecto, un co-director y tres Investigadores, un Asistente de Investigador y un becario Tesista. Con respecto a la línea de investigación comentada en el presente artículo, y tal como se indicó en el apartado anterior, integrantes del proyecto participan conjuntamente con un becario tesista en las primeras etapas del desarrollo de un aplicativo móvil para promover m-library dentro del contexto de la institución académica de la cual depende el proyecto. El resultado final será plasmado en una tesis final de carrera de grado.

Referencias

[1] ISEA S.Coop. (2009). "Mobile learning, análisis prospectivo de las potencialidades asociadas al Mobile Learning". Recuperado de http://www.iseamcc.net/eISEA/Vigilancia_tecnologica/informe_4.pdf

[2] Ramos, A.I.; Herrera, J.A; y Ramírez, M.S.; (2010), "Desarrollo de habilidades cognitivas con aprendizaje móvil: un estudio de casos", Comunicar, nº 34, v. XVII, Revista Científica de Educomunicación; ISSN: 1134-3478; pp 201-209 DOI:10.3916/C34-2010-03-20, Monterrey, México.

- [3] Villoldo, M. and Salom, A. (2012), "How to mobilize your library at low cost", *Liber quarterly*, Vol. 22 No. 2, pp.118–145.
- [4] Paterson, L. and Low, B. (2010), "Usability inspection of Digital Libraries". *ARIADNE*, Vol. <http://www.ariadne.ac.uk/issue63/paterson-low>
- [5] Smith, B. and Jacobs, M. (2010), "Libraries and patrons on the move: from bookmobiles to "m" libraries", *Reference Services Review*, Vol. 38 No. 2, <http://www.emeraldinsight.com/journals.htm?articleid=1858850&show=abstract>
- [6] Needham, G. and Ally, M. (2008), "M-libraries: Libraries on the move to provide virtual access", *Library Management*, pp.499–500, [http://scholar.google.gr/scholar?start=30&q="mobile+libraries"&hl=el&as_sdt=0,5&as_ylo=2000&as_yhi=2014#9](http://scholar.google.gr/scholar?start=30&q=)
- [7] Needham, G. and Ally, M. (2013), "M-Libraries 4, From margin to mainstream - mobile technologies transforming lives and libraries", Ed Facet Publishing, UK, ISBN 978-1-85604-944-3.
- [8] Evgenia Vassilakaki, (2014), "Mobile information services in libraries: a review of current trends in delivering information", *Interlending & Document Supply*, Vol. 42, No. 4, pp 176 - 186; DOI: <http://dx.doi.org/10.1108/ILDS-08-2014-0037>
- [9] Library Success: A Best Practices Wiki, <http://libsucces.org>
- [10] MIT Mobile Web source code <https://github.com/MIT-Mobile/MIT-Mobile-Web>
- [11] N. S. A. Karim, S. H. Darus, and R. Hussin, (2006), "Mobile Phone Applications in Academic Library Services: A Students' Feedback Survey.," *Campus-Wide Inf. Syst.*, vol. 23, no. 1, pp. 35–51, Nov. 2005.
- [12] Hennig, N., (2014), "Selecting and evaluating the Best Mobile Apps for Library Services". *Library Technology Reports*, Vol. 50, No. 8, pp 1-30.
- [13] Calvary, G. et al., (2002) The CAMELEON Reference Framework, CAMELEON Project, September, disponible en: <http://giove.isti.cnr.it/projects/cameleon/pdf/CAMELEON%20D1.1Refframework.pdf>.
- [14] Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J., A (2003), "Unifying Reference Framework for Multi-Target User Interfaces", *Interacting with Computers*, Vol. 15, No. 3, pp. 289-308.
- [15] Malizia, A., Bottoni, P., & Levialdi, S. (2010). Generating collaborative systems for digital libraries: A model-driven approach. *Information Technology and Libraries*, 29(4), 171-186. Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-78650634896&partnerID=40&md5=a76eec29569de9fb0a5bb900be8fb240>
- [16] CSS Grid Layout Module Level 1, <http://dev.w3.org/csswg/css-grid-1/>
- [17]. Indexed Database API, <http://www.w3.org/TR/IndexedDB/>
- [18] Realm Database móvil, <http://realm.io/>
- [19] CSS Flexible Box Layout Module Level 1, <http://dev.w3.org/csswg/css-flexbox/>
- [20] D. Billsus, C. A. Brunk, C. Evans, B. Gladish, and M. Pazzani, (2002), "Adaptive Interfaces for Ubiquitous Web Access," *Commun. ACM*, vol. 45, no. 5, pp. 34–38.

PROGRAMA DE I+D+I EN INGENIERÍA DE ESPACIOS VIRTUALES DE TRABAJO

Darío Rodríguez, Rodolfo Priano, Roberto García, Norberto Charczuk, Federico Ribeiro, Santiago Bianco

Laboratorio de Investigación y Desarrollo en Espacios Virtuales de Trabajo
Grupo de Ingeniería de Espacios Virtuales de Trabajo y Grupo Investigación en Sistemas de Información
Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús. Argentina. Tel +54 11 5533 5600 Ext. 5194
dariorodriguez1977@gmail.com

RESUMEN

Esta comunicación presenta los Proyectos de I&D Acreditados, las Tesis de Doctorado y las Tesis de Maestría que nuclea el Programa de I+D+I en Ingeniería de Espacios Virtuales de Trabajo; cuyas actividades de investigación están radicadas en el Grupo de Ingeniería de Espacios Virtuales de Trabajo del Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús.

Palabras clave: Espacios Virtuales de Trabajo, Proyectos I&D, tesis doctorales, tesis de Magister, Formación de Recursos Humanos.

DIRECTRIZ DEL PROGRAMA

La directriz del Programa de I+D+I en Espacios Virtuales de Trabajo es construir conocimiento aplicable a problemáticas del área sobre los procesos, las metodologías y las técnicas utilizadas para: ordenar, controlar y gestionar la tarea de mediar sincrónicamente vía tecnología web las interacciones entre personas con ubicaciones espaciales discontinuas.

El programa evoluciona en la línea de tiempo a partir de los problemas abiertos que se identifican en los proyectos y de los requerimientos que plantea el sistema productivo regional.

Este programa nuclea desde el año 2011: Proyectos de I&D Acreditados, Tesis de Doctorado y Tesis de Maestría.

PROYECTOS DE I&D ACREDITADOS

En el marco del programa se desarrollan los siguientes Proyectos de I&D Acreditados:

- UNLa 33A166: Espacios Virtuales para Trabajo Colaborativo (Código: 33A166, Periodo: 2013-2014)
- Medidas de Interacción en Espacios Virtuales para Trabajo Colaborativo (Código: 80020150200038LA, Periodo: 2016-2017)

Proyecto UNLa 33A166 Espacios Virtuales para Trabajo Colaborativo [GISI-UNLa, 2013]

Los espacios virtuales de trabajo colaborativo permiten la integración de grupos de trabajo en la que sus miembros no están físicamente contiguos. Hay una amplia literatura vinculada al modelado de las arquitecturas software que soportan este tipo de ambientes. Sin embargo, los formalismos existentes atienden la interacción entre actores y sistema y entre componentes del sistema; descuidando los aspectos de interacción humana. Este proyecto se propone desarrollar, mediante la metodología de prototipado evolutivo, los siguientes elementos: (a) herramientas para el modelado y diseño de espacios virtuales para trabajo colaborativo con énfasis en las interacciones humanas que deben soportar, (b) un arquetipo patrón de arquitectura de espacio virtual dedicados al desarrollo de proyectos grupales, y (c) herramientas de medición de interacción humana en grupos que realicen trabajo colaborativo basado en espacios virtuales.

Proyecto UNLa 80020150200038LA Medidas de Interacción en Espacios Virtuales para Trabajo Colaborativo [GISI-UNLa, 2016]

El concepto de espacio virtual para trabajo colaborativo (EVTC), surge de la fusión de los conceptos de: teletrabajo, equipos de desarrollo y espacios virtuales. Un EVTC se puede definir como un espacio basado en tecnología web que permite el trabajo colaborativo de grupos en los que sus miembros no se encuentran físicamente contiguos. Los EVTC están destinados a facilitar la mediación en el interior de equipos cuyos miembros tienen que desarrollar un objeto conceptual (por ejemplo: investigación, desarrollo de proyectos software, artículos técnicos, informes, documentación de diseño de edificios, planes de negocio, planes de inversión corporativos, entre otros). Estas actividades colaborativas, requieren que cada miembro del grupo tenga información en tiempo real de las actividades (inicio, etapa del proceso de desarrollo, culminación) que cada uno de los otros

miembros esta desarrollando en dicho espacio virtual. Por otra parte, la gestión de grupos de trabajo requiere una correcta observación y diagnóstico de la dinámica grupal. En caso de grupos de trabajo virtual mediado por la tecnología, el elemento observable se ha de ponderar y registrar por indicadores distintos de los usuales propios del entorno virtual. En este contexto, el objetivo de este proyecto es sistematizar el conocimiento existente sobre espacios virtuales de trabajo colaborativo formulando propuestas unificadas de modelado de conciencia del otro (“awareness”); y de medidas de interacción de sujetos en este tipo de espacios orientados a la gestión del grupo.

TESIS DOCTORAL

En el marco del programa se desarrollan las actividades de investigación de la siguiente Tesis Doctoral:

- Conceptualización de Espacios Virtuales para Trabajo Colaborativo (Universidad Nacional de La Plata, Finalizada 2015)

Conceptualización de Espacios Virtuales para Trabajo Colaborativo [Rodríguez, 2015]

La evolución de la calidad de las comunicaciones basadas en la tecnología web es la base de la tendencia de desarrollo de los espacios virtuales de trabajo. Como una familia emergente de aplicaciones a desarrollar, necesita herramientas para el proceso de conceptualización como entrada de los procesos de diseño. En esta tesis se proponen formalismos de modelado de interacción entre personas para Espacios Virtuales de Trabajo. Con base en estos formalismos de modelado de interacciones, se presenta un Proceso de Conceptualización de Espacios Virtuales de Trabajo orientado a cubrir las necesidades específicas de interacción, y se describe el conjunto de técnicas creadas para el desarrollo de cada tarea del proceso de conceptualización.

TESIS DE MAGISTER

En el marco del programa se desarrollan las actividades de investigación de las siguientes Tesis de Magister:

- Espacios Virtuales para la Formación de Investigadores. Elementos de Análisis y Diseño (Universidad Nacional de La Plata, Finalizada 2012)
- Modelo de Awareness Basado en Topologías de Interacción para Espacios Virtuales de Trabajo

Colaborativo (Universidad Tecnológica Nacional, Finalizada 2014)

- Proceso de Identificación de Comportamiento de Comunidades Educativas basado en Resultados Académicos (Universidad Nacional de La Plata, Finalizada 2016)
- Tecnologías Informáticas de Mediación de Interacción Aplicadas a Formación de Recursos Humanos (Universidad Tecnológica Nacional, en Desarrollo)

Espacios Virtuales para la Formación de Investigadores. Elementos de Análisis y Diseño [Rodríguez, 2012]

La realidad de nuestro país, en la que los centros de investigación en informática con capacidad de formar investigadores en todos sus niveles son escasos, lleva a la necesidad de abordar la cuestión de esquemas alternativos para la formación de investigadores. La formación mediada por tecnología surge como una posibilidad de constituir grupos de investigación en la que las instituciones universitarias con centros de investigación consolidados aporten los investigadores formados y el resto del sistema universitario las vocaciones para formarse en los procesos investigativos. También abre la posibilidad que áreas de vacancia científica sean desarrolladas con la colaboración de investigadores formados pertenecientes a Centros de I&D de otros países. En este contexto, en esta tesis busca: [a] la integración de los formalismos de modelado de dinámica de grupos con los formalismos de modelado conceptual de ambientes colaborativos; y [b] el diseño conceptual de un espacio virtual con énfasis en los procesos propios de formación de investigadores; utilizando la integración de los formalismos de modelado de dinámica de grupos con los formalismos de modelado conceptual de ambientes colaborativos.

Modelo de Awareness Basado en Topologías de Interacción para Espacios Virtuales de Trabajo Colaborativo [Herrera, 2014]

Se busca definir un modelo de awareness basado en interacciones grupales en los Espacios Virtuales de Trabajo Colaborativo (EVTC) utilizando los lineamientos establecidos en el Modelado de Interacciones en EVTC, y a partir de aquel, obtener los posibles roles y asociar la información de awareness adecuada para cada rol y topología de interacción. Esto permitirá dar un mejor soporte a la información de tipo awareness en aplicaciones colaborativas.

Proceso de Identificación de Comportamiento de Comunidades Educativas basado en Resultados Académicos [Cicliutti, 2016]

Los procesos de explotación de información se incorporan al ámbito educativo para ayudar a entender y mejorar tanto la enseñanza de los docentes como el aprendizaje de los alumnos. Entre estas cuestiones se destaca el análisis del comportamiento de comunidades educativas de forma tal de proveer al docente herramientas que ayuden a mejorar al proceso de enseñanza y aprendizaje. En este contexto, el presente trabajo de tesis tiene como objetivo proponer, estudiar y validar un proceso de explotación de información que permita identificar el comportamiento de comunidades educativas basado en resultados académicos.

Tecnologías Informáticas de Mediación de Interacción Aplicadas a Formación de Recursos Humanos [Sznajdleder, 2015]

Se pretende identificar las tecnologías de mediación de interacción aplicables a la formación de recursos humanos y definir un método que identifique la articulación del uso de estas tecnologías en los procesos formativos. Se buscará la validación del método propuesto mediante trabajo experimental en cursos universitarios. Se espera demostrar en la tesis que las tecnologías informáticas de mediación identificadas, al ser utilizadas de la manera definida en el método propuesto, permiten obtener un mejor rendimiento del proceso formativo.

FORMACIÓN DE RECURSOS HUMANOS

A la fecha de esta comunicación se han radicado en el Programa 1 tesis doctoral (1 defendida), y 4 tesis de magíster (3 defendidas). Adicionalmente, se esta trabajando en el Plan de Tesis de Magister de 2 Candidatos.

REFERENCIAS

- Cigliutti, P. 2016. *Proceso de Identificación de Comportamiento de Comunidades Educativas basado en Resultados Académicos*. Tesis de Magister en Tecnología Informática Aplicada en Educación. Facultad de Informática. Universidad Nacional de La Plata.
- GISI-UNLa, 2013. *Proyecto UNLa 33A166 Espacios Virtuales para Trabajo Colaborativo*. <http://sistemas.unla.edu.ar/sistemas/gisi/UNLa-33A166.htm>
- GISI-UNLa, 2016. *Proyecto UNLa 80020150200038LA Medidas de Interacción en Espacios Virtuales para Trabajo Colaborativo*. <http://sistemas.unla.edu.ar/sistemas/gisi/UNLa-20150200038LA.html>

Herrera, A. 2014. *Modelo de Awareness Basado en Topologías de Interacción para Espacios Virtuales de Trabajo Colaborativo*. Tesis de Magister en Ingeniería de Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional.lata.

Rodríguez, D. 2012. *Espacios Virtuales para la Formación de Investigadores*. Elementos de Analisis y Diseño. Tesis de Magister en Tecnología Informática Aplicada en Educación. Facultad de Informática. Universidad Nacional de La Plata.

Rodríguez, D. 2015. *Conceptualización de Espacios Virtuales de Trabajo*. Tesis Doctoral en Ciencias informáticas. Facultad de Informática. Universidad Nacional de La Plata.

Sznajdleder, P. 2015. *Tecnologías Informáticas de Mediación de Interacción, aplicación a la formación de Recursos Humanos*. Revista Latinoamericana de Ingeniería de Software, 3(2): 108-109, ISSN 2314-2642.

DISEÑO Y DESARROLLO DE APLICACIONES DE APRENDIZAJE UBICUO

Durán Elena, Álvarez Margarita, Unzaga Silvina, Salazar Nevelyn, Fernández Reuter Beatriz, González Gabriela, Figueroa Saritha y Cordero Ricardo

Instituto de Investigaciones en Informática y Sistemas de Información
Departamento de Informática - Facultad de Ciencias Exactas y Tecnologías
Universidad Nacional de Santiago del Estero (UNSE)
e-mail: {eduran, alvarez, sunzaga, nsalazar, bfreuter}@unse.edu.ar

RESUMEN

En la Universidad Nacional de Santiago del Estero se está llevando a cabo el desarrollo del proyecto de investigación titulado "Sistemas de información web personalizados, basados en ontologías, para soporte al aprendizaje ubicuo". El mismo tiene como objetivo principal desarrollar aplicaciones basadas en ontologías que den soporte al aprendizaje ubicuo. En este artículo se presentan los objetivos, los resultados alcanzados y la formación de recursos humanos concretada en el marco del proyecto.

Palabras clave: Aprendizaje Ubicuo, Sistemas de Información Web, Personalización, Modelo de Usuario, Ontologías.

1. DESCRIPCIÓN DEL CONTEXTO

Tradicionalmente, el aprendizaje se ha configurado como un proceso mediante el cual el aprendiz entra en contacto y adquiere conocimientos o destrezas, de alguna fuente autorizada, pero en la actualidad, las teorías más avanzadas sobre el aprendizaje sostienen que los aprendices no absorben de forma pasiva el conocimiento, personalmente significativo, sino que más bien lo crean de forma activa, a partir de su experiencia del mundo.

Por otro lado, en los últimos años, tanto Internet como la tecnología inalámbrica han registrado un gran avance permitiendo el acceso, gracias al uso de dispositivos móviles, a una gran variedad de diversa información y servicios y por lo tanto, extiende el aprendizaje haciendo realidad la idea de aprendizaje en cualquier lugar y en cualquier momento. Las personas acceden a mejores experiencias de aprendizaje en sus ambientes de la vida diaria gracias a que los contenidos educativos están disponibles, y las interacciones pueden concretarse donde los estudiantes lo necesiten, en diferentes campos de la vida diaria, sin restricción de espacio ni de tiempo (Graf y Kinshuk, 2008).

Estos avances en las teorías educativas y en las tecnologías móviles han permitido el surgimiento del aprendizaje ubicuo (u-learning), considerado como una nueva y prometedora área. Esta forma de aprendizaje designa al conjunto de actividades formativas, apoyadas en tecnología, que están accesibles en cualquier lugar y desde cualquier dispositivo (Saadiah y Kamarularifin, 2010).

El u-learning posee la característica de la personalización. Esto implica adaptar las metodologías de aprendizaje a los modos de aprendizaje de los estudiantes, de manera que se convierte en un método de aprendizaje individualizado que ayuda al estudiante más rápida y efectivamente. Es así que, la personalización en el u-learning tiene como objetivo proporcionar a los estudiantes material, actividades e información de aprendizaje adaptado y personalizado, considerando los aspectos del contexto tales como la ubicación, el momento, los diferentes dispositivos, etc.

En consecuencia, se hace necesario que estas aplicaciones de u-learning manejen información de las características diferenciadas de los usuarios, de la tecnología que se utiliza y los contextos en los que se concreta el aprendizaje. En este sentido, las ontologías son una técnica que se puede utilizar con el fin de describir dicha información y hacerla compatible para diferentes fines, accesible para agentes de software que gestionen la personalización del sistema, y lo transformen en interoperable desde distintas plataformas.

Teniendo en cuenta la problemática planteada, desde el año 2012 se está desarrollando el proyecto de investigación "Sistemas de información web personalizados, basados en ontologías, para soporte al aprendizaje ubicuo", en el Instituto de Investigaciones en Informática y Sistemas de Información, de la Facultad de Ciencias Exactas y Tecnologías, de la Universidad Nacional de Santiago del Estero. El mismo tiene como finalidad realizar propuestas metodológicas y desarrollar

aplicaciones de software, que den soporte al aprendizaje ubicuo, mejorando la interoperabilidad y usabilidad de las mismas.

En este trabajo se presentan los objetivos del proyecto, los resultados alcanzados en el marco del mismo y la formación de recursos humanos.

2. OBJETIVOS

Con este proyecto de investigación se busca realizar contribuciones teóricas y metodológicas en el campo de los sistemas de información web de apoyo al aprendizaje, por medio de la aplicación de técnicas y métodos de computación ubicua, ontologías y técnicas personalización. En consecuencia los objetivos son:

1. Diseñar modelos de entorno de aprendizaje ubicuo ajustados al contexto de la educación universitaria, a partir de un enfoque sistémico.
2. Analizar, diseñar, construir, evaluar y/o reusar ontologías para sistemas de información web de apoyo al aprendizaje ubicuo en contextos universitarios.
3. Diseñar, construir y evaluar modelos de usuario basados en ontologías para personalizar sistemas de información web de apoyo al aprendizaje ubicuo en contextos universitarios.
4. Diseñar, construir y evaluar módulos de personalización de contenidos, navegación e interfaces de usuario, basados en tecnología de agentes, para sistemas de información web de apoyo al aprendizaje ubicuo en contextos universitarios.
5. Integrar los módulos de personalización de contenidos, navegación e interfaces de usuario en sistemas de información web de apoyo al aprendizaje ubicuo en contextos universitarios.
6. Evaluar la interoperabilidad y usabilidad de los sistemas de información web personalizados, de apoyo al aprendizaje ubicuo en contextos universitarios.

3. RESULTADOS OBTENIDOS

A partir de los objetivos planteados se obtuvieron los siguientes resultados:

- a) *Modelo de entorno de aprendizaje ubicuo y aplicaciones de dicho modelo*: se diseñó un modelo de entorno de u-learning ajustado al contexto universitario a partir de un enfoque sistémico, el cual fue tomado como guía para orientar el diseño e implementación de una propuesta de capacitación docente en el uso de plataformas de aprendizaje (Figuroa et al., 2014). También, dicho modelo se aplicó, desde una perspectiva socio-cultural, como guía para

el diseño de una propuesta didáctica para la enseñanza de la asignatura Tecnología de la Matemática (Cordero y Figueroa, 2015).

- b) *Arquitectura basada en modelos ontológicos para aplicaciones de aprendizaje ubicuo*: en Durán et al. (2014a) se presentó una arquitectura para desarrollar aplicaciones informáticas de u-learning. Esta arquitectura comprende cuatro dimensiones: los modelos, las ontologías, los componentes de software y el mundo exterior. Sobre la base de esta arquitectura fue posible el desarrollo de aplicaciones que operen en entornos dinámicos de computación ubicua y se adapten a los cambios de contexto.
- c) *Representación del contexto en aplicaciones de u-learning*: se trabajó sobre la definición de un nuevo enfoque para modelar el contexto, considerando la naturaleza dinámica de su información y los servicios de personalización que se pueden ofrecer en un entorno de u-learning (González y Durán, 2015).
- d) *Modelo de dominio en sistemas de aprendizaje ubicuo*: los contenidos que un estudiante puede aprender están representados generalmente por objetos de aprendizaje. El conjunto de estos, su estructura y relaciones conforman el dominio de aprendizaje. Al respecto, en Álvarez et al. (2015a) se desarrolló el modelo conceptual del dominio con la descripción de las clases principales que lo componen, sus atributos y relaciones.
- e) *Objeto de Aprendizaje para la enseñanza de la Simulación*: se desarrollaron objetos de aprendizaje cuyo objetivo es afianzar conceptos de la Simulación como técnica experimental de resolución de problemas y que forman parte del Modelo de dominio (Salazar y Durán, 2014).
- f) *Modelo de Requerimientos de una Aplicación de Apoyo al Aprendizaje Ubicuo para el Ingreso Universitario*: en Únzaga et al. (2015) se realizó el modelado de los requisitos que permitieron validar la arquitectura mencionada anteriormente para una aplicación de u-learning en un área del curso de ingreso de la universidad.
- g) *Módulo de personalización para aplicaciones de soporte al aprendizaje ubicuo*: una de las características fundamentales del u-learning es que permite ofrecer servicios personalizados a los estudiantes. En Durán (2014) se conceptualizó las técnicas de personalización que se aplican en el u-learning y, con base en estas técnicas en Durán et al. (2014b) se presentó el módulo de personalización que es

uno de los módulos que forma parte de la dimensión componentes de software de la arquitectura. En Álvarez et al. (2015b) se describió una estrategia para personalizar recomendaciones y servicios destinados a los estudiantes basada en un enfoque de personalización híbrido. Esta estrategia permitió adaptar los contenidos, el camino de aprendizaje, la interfaz y la colaboración, tanto a las características relevantes de los estudiantes como a los aspectos que caracterizan una situación de aprendizaje particular, el ambiente donde se concreta y los dispositivos y medios que se usan. En Durán et al. (2016a) se avanzó en el desarrollo de los servicios propuestos y en el diseño de las reglas de inferencias para cada servicio.

- h) *Ontología del perfil de usuario para personalización de sistemas de u-learning universitarios*: se diseñó y construyó la ontología del perfil del estudiante para sistemas de u-learning universitarios. La misma está compuesta por las características de los estudiantes y del entorno que resultan relevantes para efectuar las tareas de personalización en dichos sistemas (González, 2014) y (González y Durán, 2014).
- i) *Modelo ontológico para personalizar aplicaciones de aprendizaje ubicuo*: en Durán et al. (2016b) se desarrolló un modelo ontológico con un esquema conceptual exhaustivo que modela la semántica del contexto de aplicaciones informáticas de u-learning y los razonamientos que se pueden hacer a partir de este modelo para inferir nuevo conocimiento con fines de personalización. Disponer de este marco de referencia permitirá a un diseñador de software tomar decisiones adecuadas con respecto a la construcción de una aplicación de u-learning, ya que contarán con todas las clases, relaciones y reglas de razonamiento.
- j) *Modelo multi-agente para recomendación personalizada de tutores de aprendizaje en tiempo real*: en Fernández Reuter et al. (2015) se propone un modelo multi-agente de recomendación de tutores que utiliza políticas configurables y permite monitorear la interacción del alumno con el entorno educativo virtual, con el fin de detectar dificultades en su aprendizaje y ofrecerle alternativas de ayuda personalizadas.
- k) *Modelo de servicios y su instanciación para una aplicación de apoyo al u-learning*: con el fin de manejar la complejidad que presentan las

aplicaciones informáticas que sirven de apoyo al aprendizaje ubicuo, se abordó el diseño de las mismas desde un enfoque basado en modelos. En Durán et al. (2016c) se presentó una nueva versión del metamodelo presentado en Durán et al. (2013) completándolo con el Modelo de Servicios. Además, se realizó la instanciación del modelo para el diseño de una aplicación de apoyo al aprendizaje ubicuo en un curso sobre Redes de Computadoras.

- l) *Diseño y construcción de una aplicación móvil personalizada de apoyo al aprendizaje de Redes de Computadoras*: se diseñó y construyó una aplicación móvil para Tablets y Smartphones con S.O. Android, que utiliza códigos QR para reconocer componentes en un laboratorio de redes de computadoras y presenta información personalizada, de esos componentes, en base al estilo de aprendizaje y nivel de conocimiento del estudiante. (Loto y Durán, 2015).
- m) *Búsqueda y Recomendación de contenido educativo*: los entornos de aprendizaje virtual poseen una gran cantidad de información útil para el desarrollo de las diferentes actividades desempeñadas por los estudiantes en las distintas etapas del proceso de aprendizaje. Sin embargo, carecen de medios de búsqueda adecuados dentro del entorno, por lo que se dificulta encontrar la información para todos aquellos estudiantes que poseen una duda puntual. En Fernández Reuter et al. (2014) se propuso un enfoque para la búsqueda y recomendación que combina técnicas de Minería de Contenido Web y Recuperación de Información, para asistir a los estudiantes en la búsqueda de información adecuada, a partir del ingreso de una consulta. Otro resultado relacionado es el presentado en Fernández Reuter y Durán (2014), donde se elaboró la búsqueda y recomendaciones para la evacuación de dudas respecto a temáticas de una asignatura, a través de una plataforma de e-learning.

4. FORMACIÓN DE RECURSOS HUMANOS

En las tablas 1 y 2 se incluye la formación de recursos humanos realizada en el marco del proyecto. Específicamente, en tabla 1 se consignan las tesis de posgrado y los trabajos finales de grado que se encuentran en desarrollo y finalizados; y en la tabla 2, los becarios con sus correspondientes trabajos y tipos de becas.

Además, con el desarrollo de este proyecto se está afianzando el grupo de investigación en las temáticas del proyecto: aprendizaje ubicuo, personalización y ontologías, lo que contribuye a

una mejora en el fondo de conocimiento disciplinar disponible, no sólo a nivel local sino también regional y nacional.

Tabla 1. Formación de Recursos Humanos: Tesis de posgrado y Trabajos Finales de Grado

TÍTULO DEL TRABAJO	ALUMNOS	NIVEL	ESTADO
Personalización en Sistemas de Aprendizaje Ubicuo	Beatriz Fernández Reuter	Tesis de Doctorado en Ciencias de la Computación, UNICEN	Tesis en ejecución.
Representación del Contexto en Aplicaciones de Aprendizaje Ubicuo.	González Gabriela	Tesis de Doctorado en Ciencias de la Computación, UNICEN	Tesis en ejecución.
QRNetwork: Aplicación móvil híbrida personalizada de apoyo al aprendizaje de Redes de Computadoras.	Matías Loto	Trabajo Final de Grado para optar por el título de Licenciado en Sistemas de Información, UNSE	Trabajo Final aprobado. Noviembre 2015.
Detección automática de roles en grupos de aprendizaje colaborativos soportados por computadora.	Gabriela Cecilia Argañaraz y Miriam Edit Calo	Trabajo Final de Grado para optar por el título de Licenciado en Sistemas de Información, UNSE	Trabajo Final aprobado, Septiembre 2015.
Sistema Multi-Agente para la personalización de Tareas en Grupos de Aprendizaje Colaborativo soportado por Computadora.	Melina Beatriz Trejo y Pablo Luis Martínez	Trabajo Final de Grado para optar por el título de Licenciado en Sistemas de Información, UNSE	Trabajo Final aprobado, Abril 2015.
Ontología del Perfil del usuario para personalización de sistemas de U-learning universitarios	González Gabriela	Trabajo Final de Grado para optar por el título de Licenciado en Sistemas de Información, UNSE	Trabajo Final aprobado. Marzo de 2014.

Tabla 2. Formación de Recursos Humanos: Becas

TIPO BECA	ORGANISMO OTORGANTE	TÍTULO DEL TRABAJO	ALUMNO	PERÍODO
Beca Interna Doctoral	Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)	Personalización en entornos de U-learning	Fernández Reuter, Beatriz	2013-2018
Beca Interna Doctoral Cofinanciada	Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) – Universidad Nacional de Santiago del Estero (UNSE)	Representación del contexto en aplicaciones de aprendizaje ubicuo	González, Gabriela	2014-2019

5. BIBLIOGRAFÍA

- Álvarez, M., Únzaga S. y Durán E. (2015a). “Modelo de dominio en sistemas de aprendizaje ubicuo”. 10° Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA. 21 y 22 de mayo de 2015. Universidad Nacional de Salta. ISBN N°: 978-987-633-133-3. Pág. 116.
- Álvarez, M., Únzaga, S. y Durán, E. (2015b). “Recomendaciones Personalizadas para Aplicaciones de Soporte al Aprendizaje Ubicuo”. XXI Congreso Argentino de Ciencias de la Computación. CACIC 2015. Junín.
- Cordero, R. y Figueroa, S. (2015). “Modelo de entorno de aprendizaje ubicuo para la enseñanza de la Tecnología de la Matemática”. X Jornadas de Ciencia y Tecnología. Argentina.
- Durán, E., Únzaga S. y Álvarez, M. (2013) “Meta-Modelo de Aplicación de Aprendizaje Ubicuo. Instanciación del Modelo de Servicios”. Anales de V Congreso Internacional de Ambientes Virtuales de Aprendizaje Adaptativos y Accesibles (CAVA 2013), ISBN: 978-950-605-762-6.
- Durán, E. (2014). “Las Técnicas de Personalización en los Sistemas de Aprendizaje Ubicuo”. Fifth International Conference on Advances in New Technologies, Interactive Interfaces and Communicability (ADNTIIC 2014), ALAIPO, Huerta Grande, Córdoba, Argentina, ISBN 978.88.96.471.37.1.
- Durán, E.; Álvarez, M. y Únzaga, S. (2014a). “Ontological model-driven architecture for ubiquitous learning applications”. EATIS 2014

- 7th Euro American Association on Telematics and Information Systems. Valparaíso (Chile). 2 al 4 de Abril de 2014.
- Durán, E.; Alvarez, M. y Unzaga, S. (2014b). "Design of a Personalization Module for U-learning application". Fifth International Conference on Advances in New Technologies, Interactive Interfaces and Communicability (ADNTIIC 2014), ALAIPO, Huerta Grande, Córdoba, Argentina, ISBN 978.88.96.471.37.1.
- Durán, E., Álvarez, M., Únzaga, S. y González, G. (2016a). "Personalization Module for U-learning Applications". Handbook of Research on Human Interaction and the Impact of Information Technologies/IGI Global. En edición.
- Durán, E., Álvarez, M., Únzaga S. (2016b). "Modelo ontológico para personalizar aplicaciones de aprendizaje ubicuo". EATIS 2016. Colombia. 27-29 de abril de 2016.
- Durán, E., Únzaga S. y Álvarez, M. (2016c). "Instanciación del Modelo de Servicios para una aplicación de apoyo al Aprendizaje Ubicuo en un curso de Redes de Computadoras". Revista TEKNOS de la Facultad de Ingeniería de la Fundación Universitaria Tecnológico Comfenalco. ISSN 1900-7388. En edición.
- Fernández Reuter, B., Durán, E. y Amandi, A. (2014). "Búsqueda y Recomendación de contenido educativo en entornos virtuales de aprendizaje". XLIII Jornadas Argentinas de Informática e Investigación Operativa. 15th Argentine Symposium on Artificial Intelligence, ASAI 2014, SADIO y Facultad de Ingeniería de la Universidad de Palermo, Buenos Aires, ISBN 1850-2784, pp. 67-74.
- Fernández Reuter, B. y Durán, E. (2014). "Framework de recomendación automática de contenidos en foros de discusión para entornos de e-learning". 7th Euro American Association on Telematic and Information Systems (EATIS 2014). Valparaíso, Chile. Proceedings published by ACM Digital Library within its International Conference Proceedings Series. ISBN 978-1-4503-2435-9. Article N°: 38.
- Fernández Reuter, B., González, G., y Durán, E. (2015). "Modelo multi-agente para recomendación personalizada de tutores de aprendizaje en tiempo real". IV Congreso Argentino de la Interacción – Persona Computador@, Telecomunicaciones, Informática e Información Científica– IPCTIIC.
- Figuroa, S; Cordero, R. Leiva, V; Aoad, V. (2014) "Modelo de entorno de aprendizaje ubicuo: su aplicación en el proceso de capacitación docente". IV Jornadas del NOA y II Jornadas nacionales de educación a distancia y tecnologías educativas. Argentina.
- González, G. (2014). "Ontología del perfil de usuario para personalización de sistemas de u-learning universitarios". XLIII Jornadas Argentinas de Informática e Investigación Operativa (43JAIIO) XVII Concurso de Trabajos Estudiantiles (EST). Facultad de Ingeniería de la Universidad de Palermo, Ciudad Autónoma de Buenos Aires. ISBN 1850-2946. Vol. 90-114. 2014.
- González, G. y Durán E. (2014). "Modelo del estudiante para sistemas de aprendizaje ubicuo: representación por medio de ontologías". IX Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2014), Universidad Nacional de Chilecito, La Rioja, ISBN 978-987-24611-1-9, pp. 298-305.
- González, G. y Durán, E. (2015). "Representación del contexto en aplicaciones de aprendizaje ubicuo". Workshop "Informática Educativa" en Seminario-Taller Internacional "Los nuevos horizontes de la Informática Educativa". Universidad Nacional de Santiago del Estero.
- Graf S. y Kinshuk (2008). "Adaptivity and Personalization in Ubiquitous Learning Systems". A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 331-338, 2008. Springer-Verlag Berlin Heidelberg.
- Loto M. y Durán E. (2015). "Diseño de una aplicación móvil personalizada de apoyo al aprendizaje de Redes de Computadoras". X Congreso de Tecnología en Educación & Educación en Tecnología. Corrientes.
- Saadiah Y., Erny A. & Kamarularifin A. J. (2010). "The definition and characteristics of ubiquitous learning: A discussion". International Journal of Education and Development using Information and Communication Technology. (IJEDICT), Vol. 6, Issue 1, pp. 117-127.
- Salazar, N. y Durán, E. (2014). "Objeto de Aprendizaje para la enseñanza de la Simulación". IX Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2014).
- Únzaga S., Álvarez M., Durán E. (2015). "Modelo de Requerimientos de una Aplicación de Apoyo al Aprendizaje Ubicuo para el Ingreso Universitario". TE&ET'15: X Congreso de Tecnología en Educación y Educación en Tecnología 2015. Argentina.

Escenarios educativos mediados por Tecnologías de la Información y la Comunicación.

Sanz Cecilia¹, Madoz Cristina¹, Gorga Gladys¹, Gonzalez Alejandro¹, Zangara Alejandra¹, Depetris Beatriz², Ibáñez Eduardo¹, Moralejo Lucrecia¹, Martorelli Sabrina^{1,3}, Artola Verónica^{1,4}, Violini Lucía¹, Salazar Mesía Natalí¹, Sanchez Mariano¹.

¹Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata

²Universidad Nacional de Tierra del Fuego

³Becaria TIPO A UNLP

⁴Becaria Doctoral CONICET

{csanz, cmadoz, ggorga, agonzalez, eibanez, l moralejo, smartorelli, vartola, lviolini, nsalazar, msanchez}@lidi.info.unlp.edu.ar,
alejandra.zangara@gmail.com, depetrisb@gmail.com

CONTEXTO

Este subproyecto forma parte del proyecto “Tecnologías para Sistemas de Software Distribuidos. Calidad En Sistemas y Procesos. Escenarios Educativos Mediados Por TIC (período 2014-2017)”, perteneciente al Instituto de Investigación en Informática LIDI, de la Facultad de Informática de la Universidad Nacional de La Plata y acreditado por el Ministerio de Educación de la Nación.

RESUMEN

En este trabajo se presentan los avances del subproyecto “Escenarios educativos mediados por Tecnologías de la Información y la Comunicación (TIC)”, que forma parte del proyecto mencionado en la sección anterior. En particular en el período 2015 e inicio de 2016.

El subproyecto tiene como objetivo general investigar, realizar desarrollos e innovar en el área de TIC y Educación. El equipo que trabaja en este subproyecto ha recorrido ya un largo trayecto en estas temáticas, y cada año se profundiza en los aportes específicos, en la formación de recursos humanos y la evolución de las líneas de investigación involucradas.

Varios de los integrantes forman parte del plantel docente y la dirección de la Especialización y Maestría en Tecnología Informática Aplicada en Educación de la Facultad de Informática de la UNLP, por lo que existe una sinergia entre el proyecto y estas carreras que permiten una retroalimentación constante.

Las principales líneas que se abordan se vinculan con el trabajo colaborativo mediado por TIC, los entornos virtuales orientados a la enseñanza y el aprendizaje, el diseño y producción de materiales

educativos digitales, y el abordaje de diferentes paradigmas de interacción persona-ordenador para enriquecer diferentes contextos y escenarios educativos.

Palabras clave: Trabajo y aprendizaje colaborativo, Entornos virtuales orientados a la enseñanza y aprendizaje, Materiales educativos digitales, Objetos de aprendizaje.

INTRODUCCION

En esta sección se presenta una revisión de antecedentes que demarcan la importancia de la investigación en la temática del proyecto y en cada una de las líneas que lo conforman.

Las teorías sobre el aprendizaje han ayudado a entender lo que es la lógica del desarrollo humano. Vygostky y Piaget en sus escritos sobre el pensamiento, el lenguaje, la maduración y el desarrollo muestran la complejidad de los procesos de aprendizaje, asociados a los mecanismos que tiene el cerebro humano para evolucionar y madurar (Rubia & Guitert, 2014).

Son varios los autores que sostienen que el aprendizaje es un proceso en donde la interacción social provee de retroalimentación, estimulación, instrucción, corrección y de andamiaje mutuo para la comprensión de significado socialmente construido (Salomon & Almog, 1998).

Por otra parte, las Tecnologías de la Información y la Comunicación se han entramado de forma tal con los procesos de enseñar y aprender, que han dado lugar a numerosas investigaciones y teorías en relación al aprendizaje con tecnología. Algunas de estas investigaciones se vinculan con el área de *Computer Supported and Collaborative Learning*, donde como indican García Valcárcel, Basilotta y López García (2014): “...se relacionan las teorías

de aprendizaje con los instrumentos tecnológicos, basado en una visión sociocultural de la cognición, que propugna la naturaleza esencialmente social de los procesos de aprendizaje y se interesa por la tecnología en cuanto al potencial que ofrece para crear, favorecer o enriquecer contextos interpersonales de aprendizaje”.

En el proyecto que aquí se presenta, se investiga sobre metodologías específicas para llevar adelante procesos de aprendizaje colaborativos mediados por TIC. Se desarrollan aplicaciones para llevar adelante dichas metodologías y se aplican en escenarios educativos concretos para analizar los resultados.

También se realizan estudios sobre las posibilidades de los entornos virtuales de enseñanza y aprendizaje. En este sentido, desde hace años se ha visto el crecimiento de sistemas de gestión de cursos (*Learning Management Systems*), con mayor cantidad de herramientas y funciones. Investigaciones más actuales indagan sobre las características de diseño que deberían tener los cursos sobre este tipo de sistemas para mejorar la *performance* de los alumnos (Jaggars & Xu, 2016).

Hay una línea de creciente auge en relación a entornos virtuales y desarrollo de cursos, que es la de los Cursos Abiertos Masivos *On-Line* (MOOC). En el trabajo de Raposo-Rivas, Martínez-Figueira y Sarmiento Campos (2015), se señala que desde sus orígenes, la mayoría de los trabajos generados en relación a los MOOC se han vinculado con su historia y concepto; en debatir los desafíos, amenazas y oportunidades; plantear estudios de caso examinando una o más plataformas y cursos y en reflexionar sobre la participación de los estudiantes en MOOC, entre otros. También afirman que luego, el foco de atención se traslada principalmente a las tasas de éxito y calidad del propio curso y que aún es necesario analizar la importancia de sus características de diseño y componentes imprescindibles.

Por otra parte, el desarrollo de materiales educativos digitales sigue siendo uno de los temas vigentes en la agenda de la investigación en el área de Tecnología y Educación. El auge de los repositorios digitales, y las herramientas de autor han permitido dar lugar a un escenario en el que los docentes producen cada vez más materiales educativos digitales. Los objetos de aprendizaje (OA) y los recursos educativos abiertos conforman parte de las temáticas de interés. Aún hace falta mayor investigación para lograr que estos materiales puedan ser reutilizados efectivamente por los investigadores. Así diversos temas como sistemas recomendadores de objetos de aprendizaje, ensamblaje de OA para la conformación de itinerarios de aprendizaje, estandarización de metadatos para materiales educativos digitales, y

modelos de empaquetamiento, entre otros, se han vuelto parte del quehacer de los investigadores del área.

Es por esto que se continúa con esta línea de investigación y se han logrado avances en este sentido.

Estos son algunas de las líneas que se abordan desde el proyecto presentado, que se mencionan de manera más sintética en la siguiente sección.

LINEAS DE INVESTIGACION / DESARROLLO

Se mencionan aquí las principales líneas de investigación y desarrollo abordadas en el marco del proyecto:

- Entornos virtuales de enseñanza y aprendizaje. Proyectos que los incorporan, metodologías, desarrollos, evaluación de su calidad y experiencias.
- Materiales educativos digitales. Objetos de aprendizaje. Mutimedias e hipermedias.
- Formación de recursos humanos en el uso de TICs.
- Trabajo colaborativo mediado por TICs. Conceptualización, análisis y desarrollo de software y metodologías.
- Simuladores, laboratorios virtuales y remotos. Ambientes virtuales 3D. Redes sociales. Aplicaciones, desarrollos, experiencias.
- Dispositivos móviles como soporte para el trabajo colaborativo y para prácticas de aula extendida y blended- learning.

RESULTADOS ESPERADOS/OBTENIDOS

Se detallan a continuación algunos de los resultados obtenidos en este proyecto, para algunas de las líneas de I/D mencionadas que han tenido mayor desarrollo durante el año 2015.

- **Entornos virtuales de Enseñanza y Aprendizaje (EVEA). Campus virtuales. Entornos inmersivos y Redes sociales.**

Se continúa realizando un taller anual sobre entornos virtuales de enseñanza y aprendizaje, en el marco de la Maestría de Tecnología Informática Aplicada a Educación. Se estudian tendencias sobre este tipo de sistemas y se realizan análisis de casos. Se está indagando sobre herramientas de foros con posibilidad de intercambio de *podcasts*. En este sentido se ha iniciado la realización de un trabajo de grado vinculado a este tema.

Se está terminando un trabajo de grado sobre “Desarrollo de una plantilla para Moodle que

orienta a los docentes en la incorporación de herramientas de la Web 2.0” (Romanut, Gonzalez, Madoz, 2014).

Al mismo tiempo, se ha iniciado un proyecto de actualización del actual entorno virtual WebUNLP, con avances basados en las investigaciones de años previos sobre EVEAs.

También, se continúa investigando sobre la integración de estos entornos con repositorios de Objetos de Aprendizaje para la implementación de cursos abiertos masivos on-line (MOOC).

Respecto de la investigación sobre mundos virtuales y entornos inmersivos, se ha finalizado una tesis de maestría vinculada a las posibilidades y barreras de *Second Life* para el escenario educativo. En este marco se ha desarrollado un estudio de caso en *Second Life*, con participación de docentes de diferentes universidades (Escobar, Sanz & Zangara 2015) y se ha elaborado como aporte una guía de competencias requeridas por los docentes para el uso de este tipo de entornos. Esta tesis es dirigida por dos miembros del proyecto. Como continuidad de este trabajo se está dirigiendo otra tesis de maestría vinculada al uso de *Second Life* para el desarrollo de sub-competencias lingüísticas y discursivas en inglés. La tesis será presentada durante 2016 (Quintín, Sanz & Zangara, 2014).

Durante el 2015, se ha obtenido la Re-certificación del sistema de gestión de calidad del “Diseño y Realización del curso de nivelación a distancia para el pre-ingreso a la Facultad de Informática” (Norma IRAM-ISO 9001:2008). También se ha continuado con el desarrollo del sistema WebECALEAD vinculado a la evaluación de procesos educativos mediados por TIC.

Se ha finalizado durante 2015 una tesis de maestría, dirigida por un miembro del proyecto, vinculada a modelos de integración de Tecnologías de la Información y la Comunicación en Instituciones de Educación Superior. En particular, se ha desarrollado un modelo para evaluar el grado de integración de TIC, y se ha realizado un caso de estudio en el que se analiza el grado de utilización de TIC, en general, y de EVEA en particular, en la Facultad de Ciencias Económicas de La Plata (Nóbile & Sanz, 2015).

Se finalizó otra tesis de maestría, dirigida por un miembro del proyecto, sobre evaluación en entornos virtuales (Rainolter, Zangara, Garmendia & Malvassi, 2015).

▪ **Diseño, producción y evaluación de Materiales educativos digitales. Objetos de aprendizaje y sus Repositorios. Simuladores. Laboratorios Remotos y Virtuales.**

Respecto de esta línea del proyecto se han realizado diferentes avances durante el 2015, que se detallan a continuación.

Se ha avanzado en la aplicación de un simulador que permite introducir los conceptos iniciales de la programación concurrente a los alumnos de primer año de las carreras de informática (De Giusti, 2015). Se ha iniciado el desarrollo de una tesis doctoral que propone la investigación y desarrollo de *frameworks* para la creación de objetos de aprendizaje (Violini, Pesado & Sanz, 2016).

Se ha finalizado una tesis de maestría, dirigida por un miembro del proyecto, relacionada con el desarrollo de un material educativo hipermedial para la enseñanza de la lectura pianística (Tuñez & Gonzalez, 2015). También se finalizó otra tesis de maestría sobre el desarrollo de e-actividades, siguiendo las conceptualizaciones y metodología planteadas en el marco de este subproyecto sobre el tema de e-actividades. La tesis fue dirigida por un miembro del proyecto (Oltolina & Zangara, 2015).

Se está avanzando en una tesis de maestría en la que se plantea una metodología para evaluar el uso de materiales educativos hipermediales en el aula para la enseñanza de la matemática en carreras de ingeniería (Del Río, Bucari & Sanz, 2015).

Se ha finalizado un trabajo de especialización en Tecnología Informática Aplicada en Educación, vinculado al estudio de microscopios virtuales para la enseñanza y aprendizaje de disciplinas que requieren este tipo de herramientas. Se han realizado publicaciones al respecto (Martorelli, Martorelli, Sanz & Pesado, 2015). Se ha trabajado en el diseño de una aplicación educativa para la televisión digital interactiva, orientada a la identificación de preparados virtuales de tejido animal (Martorelli, Martorelli, Sanz, Abásolo & Pesado, 2015).

Durante 2015 se ha creado un sitio específico de la metodología (CROA) desarrollada en el marco de este proyecto para la Creación de Objetos de Aprendizaje (Sanz, Barranquero & Moralejo, 2015), y se ha utilizado en el marco de un concurso de desarrollo de OA en la Facultad de Informática (De Giusti et al., 2015).

Se ha finalizado un trabajo de grado relacionado con la realización de materiales educativos basados en Realidad Aumentada (Salazar, Gorga & Sanz 2015). También se ha avanzado sobre la evaluación de la motivación intrínseca de los estudiantes en la utilización de este tipo de materiales educativos.

Se ha finalizado una tesis de maestría, dirigida por un miembro del proyecto, vinculada al desarrollo de un marco que permita comparar metodologías de creación de objetos de aprendizaje (Maldonado, Sanz & Pampillón-Fernandez, 2015). Se ha entregado para su evaluación una tesis de maestría sobre sistemas recomendadores y ensambladores de

OA (Astudillo, Sanz & Santacruz Valencia, 2014). La dirección de estas dos tesis ha estado a cargo de un miembro del proyecto.

Asimismo, continúa en proceso de desarrollo un trabajo de especialización en la que se presenta un modelo conceptual para relevar el estado de arte de materiales digitales para la enseñanza de español como segunda lengua en niños de 5 a 7 años (Balanta, 2015).

▪ **Trabajo colaborativo mediado por TICs.**

En el marco de esta línea de investigación se ha avanzado en varios sentidos, se ha profundizado en el trabajo sobre metodologías específicas para el trabajo colaborativo mediado por TIC, así como también en su proceso de evaluación. Se detallan a continuación los avances.

Durante el año 2015 se ha desarrollado una experiencia de trabajo colaborativo en línea para el aprendizaje de temas de programación, siguiendo una estrategia de aprendizaje basado en problemas. Se han obtenido resultados positivos que fueron presentados en dos congresos relacionados a la temática (Gonzalez, Madoz, Depetris & Aguilera Mallea, 2015).

Se está finalizando una tesis de doctorado que se focaliza en el uso de dispositivos móviles para el aprendizaje, y se han implementado una serie de juegos cooperativos y colaborativos, utilizando el marco de diseño MADE-mLearn. Se han concretado experiencias ad-hoc para el testeo y análisis de posibilidades de dichos juegos. La tesis se entregará durante el 2016. Al mismo tiempo uno de los miembros del proyecto y que dirige esta tesis doctoral, es asesor en estas temáticas en un proyecto vinculado de la Universidad Nacional de Santiago del Estero.

También, se está finalizando una tesis de doctorado sobre “Interacción e interactividad en el trabajo colaborativo mediado por tecnología digital”, realizada por uno de los miembros del proyecto y dirigida por otro. Durante 2014 y 2015 se han realizado sesiones de seguimiento de procesos colaborativos mediados por diferentes tecnologías, y para ello se ha definido una metodología para la evaluación de dichos procesos. Estos avances se han ido presentando en diferentes publicaciones en congresos (Zangara & Sanz, 2015).

Se finalizó una tesis de maestría co-dirigida por uno de los miembros del proyecto en la que se desarrolló una matriz de valoración para analizar el proceso de resolución de problemas, con indicadores que permitieron reconocer y caracterizar las situaciones surgidas del aprendizaje colaborativo virtual, manifestadas a través del

discurso escrito en los foros grupales (Bacino, Massa & Zangara, 2015).

Se ha finalizado un trabajo de Especialización en TIAE, vinculado a la utilización de dispositivos móviles para el aprendizaje colaborativo (Gallo, Thomas & Sanz, 2015).

En cuanto a los proyectos vinculados con la temática y acuerdos de cooperación, el III- LIDI participa en los siguientes:

- Se ha firmado un acuerdo de colaboración en estos temas con la Universidad de Zaragoza. En este marco una docente de dicha Universidad ha realizado una estada de investigación en el III LIDI durante el 2015.

- Se participa en la Red constituida por universidades de Iberoamérica en el marco del programa "Pablo Neruda" dentro del Espacio Iberoamericano del Conocimiento (EIC) y de la Organización de los Estados Iberoamericanos (OEI) orientada a la movilidad de estudiantes y docentes de doctorado. Período: 08/2009 a 07/2010-Renovación, en el período: 08/2010 a 12/2014. En este ámbito se ha finalizado en 2015 un libro con 3 capítulos relacionados con temas de investigación del proyecto y 3 capítulos vinculados a las investigaciones abordadas por otro grupo de México que participa de la red. El libro se encuentra en editorial.

- Dos miembros del proyecto han participado en el proyecto: “*Replikants: Towards a New Generation of Human-like Agents*”, con el N° TIN201124660, del Programa Tecnología de la Información y las Comunicaciones de la Subdirección General de Proyectos de Investigación del Ministerio de Economía y Competitividad del gobierno de España, que se ha desarrollado durante el período comprendido desde septiembre de 2011 a septiembre de 2015.

- RedAUTI: Red temática en Aplicaciones y Usabilidad de la Televisión digital Interactiva. En este ámbito se analizan materiales educativos para la TVDI. Se ha trabajado en el tema de MOOC en relación a la TVDI.

FORMACION DE RECURSOS HUMANOS

Cuatro integrantes de esta línea de investigación están desarrollando su Tesis de Magister en Tecnología Informática Aplicada en Educación.

Uno de los miembros se encuentra finalizando una beca Tipo A de la UNLP. Otro miembro tiene desde 2015 una beca Tipo A de la UNLP.

Se dirigen Tesis de Magister, Trabajos de Especialista y Tesinas de Licenciatura de la Facultad de Informática en temas relacionados con el Proyecto.

En 2015, se han aprobado 7 trabajos de Maestría y 1 de Especialización en el área de Tecnología y Educación, dirigidos por miembros de este proyecto.

Hay 1 trabajo de Maestría y otro de especialización terminados y esperando su evaluación, todos dirigidos por miembros del proyecto, otros 6 trabajos de maestría vinculados al proyecto en desarrollo con propuestas aprobadas, 1 trabajo de Doctorado en desarrollo sobre temas de *m-learning* y 1 de doctorado relacionado con trabajo colaborativo mediado por TIC. Finalmente, se están iniciando 2 trabajos finales de Especialización vinculados al proyecto. Además, hay dos trabajos de grado en desarrollo.

BIBLIOGRAFIA

- Astudillo, G., Sanz, C. & Santacruz Valencia, L. (2014). Sistemas ensambladores de objetos de aprendizaje. IX Congreso sobre Tecnología en Educación & Educación en Tecnología (La Rioja, 2014). Organizado por la Red de Universidades con Carreras en Informática (RedUNCI). ISBN: 978-987-24611-1-9. 373-382. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/38680>
- Bacino, G., Massa, S. & Zangara, A. (2015). Aula extendida en la educación superior en Ingeniería. Una propuesta de aplicación en el área tecnológica básica de electrotecnia. Maestría en TIAE. Aprobada en 2015. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/45476>
- Balanta, A., Zangara A. “Diseño, desarrollo y evaluación de material multimedia interactivo para la enseñanza de español como lengua extranjera a niños”. Propuesta de trabajo Especialización en TIAE. 2014.
- De Giusti, A., Zangara, A., Sanz, C., Moralejo, L., Barranquero, F. & Naiouf, M. (2015). Producción de objetos de aprendizaje para la enseñanza universitaria. Convocatoria a los docentes de la Facultad de Informática de la UNLP. XXI Congreso Argentino de Ciencias de la Computación. Junín, 2015. ISBN: 978-987-3806-05-6. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/50642>
- De Giusti, L., Leibovich, F., Chichizola, F., Naiouf, M. & De Giusti, A. (2015). Incorporando conceptos en la enseñanza de Concurrency y Paralelismo utilizando el entorno CMRE? IV Workshop de Innovación en Educación en Informática. CACIC 2015. Junín, Buenos Aires, Argentina.. ISBN: 978-987-3806-05-6. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/50645>
- Del Río, L., Bucari, N. & Sanz, C. (2015). Material Didáctico Hipermedia para la Enseñanza de la Matemática en Carreras de Ingeniería: Inicios de una Investigación. Educación Matemática en Carreras de Ingeniería. XIX Encuentro Nacional. Universidad Tecnológica Nacional, Sede San Nicolás. Disponible en: <http://www.frsn.utn.edu.ar/EMCI/files/Resumes.pdf>
- Escobar Gutierrez, M., Sanz, C., Zangara, A. Posibilidades educativas del entorno 3D Second Life para docentes. Estudio de caso con docentes de un postgrado de la Universidad Nacional de la Plata. Tesis de Maestría en TIAE. Aprobada en 2015. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/49862>
- Gallo, S., Thomas, P. & Sanz, C. (2015) Dispositivos móviles en actividades educativas colaborativas: análisis y recopilación de experiencias. Trabajo de Especialización en TIAE. Aprobado en 2015. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/49925>
- García-Valcárcel, A., Basilotta, V & López García, C. (2014). Las TIC en el aprendizaje colaborativo en el aula de Primaria y Secundaria. Comunicar, n° 42, v. XXI, , Revista Científica de Educomunicación; ISSN: 1134-3478, 65-74
- Gonzalez, A., Madoz, C., Depetris, B. & Aguil Mallea, D. (2015). Una propuesta de trabajo colaborativo en línea para el desarrollo de algoritmos y programas. X Congreso sobre Tecnología en Educación & Educación en Tecnología (TE & ET). Corrientes, 2015. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/48362>
- Gonzalez, A., Madoz, C., Depetris, B. & Aguil Mallea, D. (2015). Formación de docentes y alumnos en el desarrollo de actividades de programación colaborativa en línea. XXI Congreso Argentino de Ciencias de la Computación. Junín, 2015. ISBN: 978-987-3806-05-6. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/50627>
- Jaggars, S. & Xu, D. (2016). How do online course design features influence student performance? Computers & Education. Volume 95, 270-284.
- Maldonado Mahauad, J., Sanz, C. & Pampillón-Fernandez, A. (2015) Desarrollo de un marco de análisis para la selección de metodologías de diseño de objetos de aprendizaje (OA) basado en criterios de calidad para contextos educativos específicos. Tesis de Maestría en TIAE. Aprobada en 2015.

- Disponible en: <http://sedici.unlp.edu.ar/handle/10915/45063>
- Martorelli, S., Martorelli S.R., Sanz, C. & Pesado, P. (2015). Revisión de microscopios virtuales: en busca de características que favorezcan el trabajo colaborativo. XXI Congreso Argentino de Ciencias de la Computación. Junín, 2015. ISBN: 978-987-3806-05-6. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/50272>
 - Martorelli, S., Martorelli, S.R., Sanz, C., Abásolo, M.J. & Pesado, P. (2015). HistoTV: Primeros pasos hacia una aplicación educativa para TVDi sobre Histología Animal. 2015 International Conference on Interactive Digital Television. JAUTI 2015 IV Iberoamerican Conference on Applications and Usability of Interactive TV. Palma de Mallorca, España.
 - Nóbile, C. & Sanz, C. Procesos de Integración de Tecnologías de la Información y Comunicación en instituciones de educación superior. El caso de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata. Tesis de Maestría en TIAE. Aprobada en 2015. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/45725>
 - Oltolina Giordano, M. & Zangara, A. (2015). La formación de competencias digitales de estudiantes de profesorado universitarios: la estrategia de e-actividades en un modelo de aula extendida. Tesis de Maestría en TIAE. Aprobada en 2015. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/49524>
 - Quintín, E., Sanz, C. & Zangara, A. (2015). La influencia de los juegos de rol en Second Life en el desarrollo de las competencias lingüística y discursiva en inglés a nivel oral. Propuesta de tesis aprobada en 2015. Facultad de Informática. UNLP.
 - Rainolter, A., Zangara, A., Garmendia, A. & Malvassi, S. (2015). Aportes de la tecnología informática para la evaluación de los aprendizajes en el ámbito de la enseñanza universitaria en entornos virtuales. El caso del Sistema de Educación Abierta y a Distancia (SEAD) de la Universidad Nacional de Mar del Plata (UNMdP). Tesis de Maestría en TIAE. Aprobada en 2015. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/47533>
 - Raposo-Rivas, M., Martínez-Figueira, E. y Sarmiento Campos, J. (2015). Un estudio sobre los componentes pedagógicos de los cursos online masivos. Comunicar, n° 44, v. XXII, 2015 | Revista Científica de Educomunicación | ISSN: 1134-3478; e-ISSN: 1988-3293.
 - Romanut, L., Gonzalez, A., Madoz C. Desarrollo de una plantilla para Moodle que oriente a los docentes en la incorporación de herramientas de la Web 2.0". Propuesta de trabajo de grado presentada durante 2014.
 - Rubia, B. & Guitert, M. (2014) ¿La revolución de la enseñanza? El aprendizaje colaborativo en entornos virtuales (CSCL). Revista Comunicar. Número 42. V. XXI. DOI: <http://dx.doi.org/10.3916/C42-2014-a2>
 - Salazar Mesía, N., Gorga, G. & Sanz, C. (2015). EPRA: herramienta para la enseñanza de conceptos básicos de programación utilizando realidad aumentada. X Congreso sobre Tecnología en Educación & Educación en Tecnología. Corrientes, 2015. ISBN: 978-950-656-154-3, 426-435. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/49071>
 - Salazar Mesía, N., Gorga, G. & Sanz, C. (2015). Plan de evaluación del material educativo digital EPRA. Propuesta de indagación sobre la motivación intrínseca. XXI Congreso Argentino de Ciencias de la Computación. Junín, 2015. ISBN: 978-987-3806-05-6. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/50310>
 - Salomon, G., & Almog, T. (1998). Educational psychology and technology: A matter of reciprocal relations. Teachers' College Record, 100, 1, 1-20.
 - Sanz, C., Barranquero, F. & Moralejo, L. (2015). CROA. Disponible en: <http://croa.info.unlp.edu.ar>
 - Tuñez, M., Shifres, F. & Gonzalez A. (2015). Lectura pianística: Diseño de material educativo en hipermedia para el desarrollo del proceso lector de partituras pianísticas. Tesis de Maestría en TIAE. Aprobada en 2015. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/47926>
 - Violini, L., Sanz, C. & Pesado, P. (2015). "Metodologías y frameworks para el diseño y la creación de objetos de aprendizaje". Propuesta de tesis doctoral aprobada en febrero de 2016.
 - Zangara A. & Sanz, C. (2015). Collaborative learning work. Evaluation model using the ethnographic methodology of "analysis of social networks". 7th International Conference on Education and New Learning Technologies. ISBN: 978-84-606-8243-1. 5373-5380 Barcelona, España.

Computación Aplicada: Búsqueda y Desarrollo de Nuevas Estrategias de Enseñanza y Recursos TIC

**Adair Martins, Carina Fracchia, Claudia Allan, Susana Parra,
Renzo García, Estefania Zurbrigk, Natalia Baeza, Lorena Robles, Daniel Benilla,
Ana Alonso de Armiño, Roberto Laurent**

Departamento de Computación Aplicada / Facultad de Informática
Departamento de Electrotecnia / Facultad de Ingeniería
Universidad Nacional del Comahue

Dirección: Buenos Aires 1400, 8300 – Neuquén

Teléfono: 0299 - 4490300 int. 429

e-mails: {adair.martins, carina.fracchia, claudia.allan, susana.parra}@fi.uncoma.edu. ar
{rgarcia.inf, baeza.natalia, anacarolinaalonsode}@gmail.com,
{dymbe2, dopanga, zureste}@hotmail.com, {rlaurent@yahoo.com.ar}

Resumen

Resultados de experiencias de incorporación de nuevas tecnologías en el aula nos han permitido observar un aumento del interés de los estudiantes, favoreciendo en ellos la adquisición de nuevas destrezas. También se han visto fortalecidos los procesos colaborativos de enseñanza y aprendizaje en los distintos ámbitos de educación. En la búsqueda del acercamiento de la universidad al medio se continúa avanzando en el desarrollo de herramientas computacionales basadas en software libre y en la utilización de aplicaciones de objetos de aprendizaje y de realidad aumentada. En este trabajo se presentan algunos resultados alcanzados en las líneas de investigación.

Palabras clave: Computación aplicada, TIC, educación, realidad aumentada, objetos de aprendizaje.

Contexto

Las líneas de investigación y desarrollo que se vienen trabajando están enmarcadas en el proyecto “Simulación y Métodos Computacionales en Ciencias y Educación”, Facultad de Informática (FAIF), Universidad Nacional del Comahue (UNCo), aprobado en el año 2014. Participan docentes, estudiantes de la carrera de Licenciatura en Ciencias de la Computación de la FAIF. Colaboran como asesores, docentes de la Facultad de Ingeniería (FI), UNCo, y de la Universidad de Brasilia (UCB), Brasil.

Introducción

En la actualidad las Tecnologías de la Información y Comunicación (TIC) están presentes en el día a día en todas las áreas del conocimiento y principalmente con significativas aplicaciones en las áreas de educación y de las ciencias en general. Su utilización e incorporación en el sistema

educativo es indispensable para la generación de una oferta educativa de calidad.

Muchos de los recursos tradicionales utilizados en la enseñanza primaria, por ejemplo el libro, donde los contenidos abordados están relacionados con la representación tridimensional o la espacial, crean un filtro cognitivo debido a que los objetos (diagramas/órganos del cuerpo humano/elementos del relieve) son representados en dos dimensiones.

En base a lo mencionado se continúan con los estudios en la búsqueda y en la utilización de aplicaciones de realidad aumentada (RA). Esta tecnología combina el mundo real con el virtual, posee la capacidad de enriquecer elementos de la realidad con información detallada. Viene siendo utilizada en distintos ámbitos, principalmente en el ámbito educativo complementando los materiales didácticos con modelos virtuales que estimulen la percepción y ayuden a la comprensión de los conceptos [1,2].

En las últimas décadas, se viene investigando nuevas formas de comunicación e interacción con la computadora a través del diseño de Objetos de Aprendizaje (OA) para la organización de contenido educativo y con el objetivo de propiciar el aprendizaje. Para el proceso de implementación de un OA se adecuan los contenidos y actividades para su presentación en un entorno virtual, se completan los metadatos y se genera el paquete SCORM para almacenarlo en un repositorio y posteriormente publicarlo. [3,9].

En la actualidad es posible encontrar modelos que hacen uso de simulación en las áreas de informática, ingeniería,

economía, etc. La simulación es una técnica con la cual se desarrolla un modelo de un sistema real y mediante experimentos con el mismo permite entender su comportamiento. Tiene aplicación en estudios de sistemas complejos sobre los que resulta costoso, difícil o peligroso llevar a cabo estudios reales. La utilización de las técnicas de simulación mediante computadoras, sumado a los actuales entornos gráficos e interfaces cada vez más amigables ofrecen a la enseñanza un aporte muy valioso [4,12].

La internet ha traído grandes beneficios, pero con el diseño de páginas Web no accesibles, ha aumentado la exclusión de personas con algún tipo de discapacidad. Cuando se habla de accesibilidad Web se refiere a la posibilidad de que todas las personas, independientemente de su capacidad, puedan navegar e interactuar con la Web sin dificultades. Algunos integrantes del grupo en conjunto con docentes de la UNCo y profesores de nivel medio, participaron en el diseño de sitios de la UNCo considerando los puntos relacionados con la accesibilidad Web [10,11].

Líneas de Investigación, Desarrollo e Innovación

El grupo viene trabajando en las siguientes líneas:

- Desarrollo y Uso de TIC
- Estudio de Objetos de Aprendizajes y sistemas para la creación de repositorios
- Estudio y análisis de los aspectos tecnológicos contemplados en el trabajo con Realidad Aumentada
- Simulación de modelos
- Accesibilidad Web

Resultados y Objetivos

Se continúan realizando experiencias utilizando la metodología de RA con alumnos y docentes de escuelas primarias de la ciudad de Neuquén. Esto ha permitido observar que el trabajo con RA ha potenciado especialmente la adquisición de una variedad de habilidades tales como la capacidad espacial, la investigación y fundamentalmente la comprensión conceptual. La incorporación de estas nuevas tecnologías, en el aula favorece despertar el interés de los estudiantes, convirtiendo la materia a tratar en algo más que mera información retransmitida. Transforma el aprendizaje en un juego, donde los nuevos descubrimientos permiten adquirir destrezas, incrementar la confianza y la responsabilidad [2].

En nuestra investigación, la búsqueda de objetos 3D de uso libre no resultó ser una tarea trivial, teniendo en cuenta que estos recursos constituyen la parte fundamental de la tecnología RA. Se encontró repositorios de objetos 3D, de elementos multimediales (vídeos, audio, imágenes, etc.), aunque en muchas ocasiones estos no son de calidad o no representan justamente los recursos que se necesita para la aplicación o trabajo que se está desarrollando. Por este motivo uno de los objetivos del proyecto de investigación es crear un repositorio con los objetos que se vayan desarrollando y dejar disponibles para su uso principalmente en el ámbito educativo. Si bien existen herramientas de autoría, muchas son pagas y las que ofrecen versiones gratuitas son por lo general con funcionalidades muy limitadas. Esto nos abre un camino en busca de lograr la creación de nuevos entornos de RA que se puedan emplear para personalizar las experiencias de aprendizaje de cualquier estudiante de

una forma autónoma. Los resultados obtenidos en las experiencias realizadas en las escuelas primarias, donde se han trabajado contenidos de ciencias naturales y sociales, tales como sistemas del cuerpo humano, relieve y fenómenos naturales, entre otros, han permitido observar las bondades que ofrece una tecnología como lo es RA en cuanto a motivación y mejora de la comprensión de los temas, como así también la necesidad de formar al profesorado en nuevas tecnologías para su uso en el aula, en particular la capacitación en el uso y diseños de recursos educativos empleando RA [1, 5].

De acuerdo a lo mencionado se ha dado origen a un proyecto de extensión que se desarrollará en el transcurso del presente año, denominado “Realidad Aumentada: Recurso Educativo para la Enseñanza en la Educación Primaria”, conformado por docentes y estudiantes de la Facultad de Informática de la UNCo y docentes del nivel primario de diferentes escuelas públicas y de gestión privada de Neuquén.

Se puede mencionar la colaboración de algunos integrantes del grupo de investigación en proyectos de la FAIF. El proyecto: Entrenamiento en la Programación de la Computadora a partir de una aplicación para competencias y Olimpiadas de Programación para Escuela Media desarrollados en los años 2014 y 2015 respectivamente. Al anterior se suma la participación en los IX Juegos Olímpicos del Comahue 2015 donde se estudia el aprendizaje competitivo como estrategia de aprendizaje de programación. Esto ha permitido avanzar en la investigación realizada en la línea de Desarrollo y Uso de TIC dando origen a la publicación [6]. Siguiendo en la temática se ha avanzado en el desarrollo de una herramienta computacional,

basada en software libre, que contempla la edición individual o colaborativa de algoritmos. El objetivo principal es que la misma pueda ser incorporada en los cursos iniciales de programación de la FAIF para mejorar el rendimiento académico de los estudiantes. Una observación importante que se puede mencionar es que pedimos a los estudiantes que sean creativos en sus soluciones y como docentes debemos procurar desarrollar nuestra creatividad en pos de una mejora de nuestras prácticas en la docencia. En este sentido se busca participar de espacios que promuevan el intercambio de ideas y conocimientos entre investigadores y profesionales de distintas disciplinas implicados en el desarrollo de la creatividad en la investigación y la educación [7].

Se continúa avanzando en el estudio y en la utilización de Objeto de Aprendizaje (OA). Se han desarrollado diferentes OA los cuales han sido incorporados en el dictado de las asignaturas Cálculo Numérico y Métodos Computacionales para el Cálculo del Departamento de Computación Aplicada de la Facultad de Informática de la UNCo. En nuestra experiencia se eligió una herramienta de software libre, que contaba con interface intuitiva y amigable. Los resultados obtenidos mostraron gran aceptación y mejora en la motivación de los estudiantes. El mismo dio origen a la publicación [8].

Continuando con la investigación sobre la accesibilidad Web se ha realizado en años anteriores el análisis del sitio Web de la UNCo y de la Plataforma de Educación a Distancia (PEDCo) haciendo uso de herramientas gratuitas que cumplen con los estándares del W3C (World Wide Web Consortium). El mismo fue realizado con herramientas

que evalúan el grado de accesibilidad de páginas Web y ayudan a identificar los aspectos que se deberían corregir para lograr un mayor grado de accesibilidad y usabilidad. Los resultados obtenidos fueron poco satisfactorios y se propusieron modificaciones para mejorar el grado de accesibilidad del sitio Web de la Universidad [10].

Recientemente se han efectuado estudios similares a los previamente realizados, del sitio Web de la UNCo y de la PEDCo. Una de las herramientas empleadas para este análisis fue la herramienta HERA, la cual proporciona información acerca de los elementos a verificar manualmente e indica cómo debe realizarse dicha verificación. Luego de analizar una página presenta dos vistas de la misma, una en modo gráfico y otra del código HTML, donde se destacan los puntos en cuestión con colores distintivos. La otra herramienta utilizada fue la Examinador. La misma presenta un informe detallando los resultados del análisis y les coloca un puntaje de 1 a 10, destacando los elementos que se deben revisar. Se puede mencionar que los resultados obtenidos en comparación con los obtenidos en los análisis anteriores, que en el caso de la página de la UNCo, la puntuación ha subido de 1.9 a 4.9. En el caso de la PEDCo, también la puntuación ha subido de 3.8 a 7.4. Sin dudas, aunque se ha mejorado los niveles de accesibilidad en el sitio de la Universidad, se requiere que continúen los esfuerzos en ese sentido para alcanzar niveles aún más altos.

En la línea de simulación y modelos se sigue investigando la aplicación de distintos software libre para la simulación de redes de computadoras y se continúa trabajando en el análisis y en el diseño de un algoritmo para el desarrollo de un

módulo que permita realizar la simulación del potencial eólico, para la estimación de la energía anual, de una determinada localidad.

Formación de Recursos Humanos

El proyecto posee doce integrantes. Participan docentes, estudiantes de la carrera de Licenciatura en Ciencias de la Computación, asesores locales y del exterior. Una integrante finalizó los cursos del doctorado en: “Enseñanza de las Ciencias Exactas y Naturales” en la Facultad de Ingeniería (FI), UNCo y se encuentra en etapa de tesis. Otra docente finalizó los cursos de la maestría “Enseñanza de las Ciencias Exactas y Naturales”, FI, UNCo y se encuentra en etapa de tesis. Una tercera integrante está actualmente realizando los dos últimos cursos de los cinco obligatorios de la maestría en Tecnología Informática Aplicada a la Educación en la Universidad Nacional de La Plata. Una estudiante se encuentra próxima a finalizar su tesis de grado.

Referencias

- [1] A. Martins, C. Fracchia, C. Allan, S. Parra, R. García, E. Zurbrigk, N. Baeza, L. Robles, D. Benilla, A. Alonso de Armiño, R. Laurent. TIC y Métodos Computacionales en el Ámbito Educativo, XVII Workshop de Investigadores en Ciencias de la Computación (Wicc 2015). Salta, 2015.
- [2] J. Cubillo Arribas, S. Martín Gutiérrez, M. Castro Gil, A. Colmenar Santos. Recursos digitales autónomos mediante realidad aumentada. RIED. Revista Iberoamericana de Educación a Distancia, vol.17, N°2, pp. 241-274, 2014.
- [3] L. Weitzel, P. Quaresma, R. Britos, R. Pimentel, Recuperación de Objetos de Aprendizaje Accesibles, X Conferência Latino-Americana de Objetos e Tecnologías de Aprendizagem (LACLO 2015), Maceió, 2015.
- [4] Sokolowski, J. A., Banks C. M., Handbook of Real-World Applications in Modeling and Simulation, Wiley, ISBN: 978-1-118-11777-4, 2012.
- [5] C. Fracchia, A. Alonso de Armiño, A. Martins, Realidad aumentada aplicada a la enseñanza de Ciencias Naturales. Revista: TE & ET; N°. 16. ISSN: 1850-9959. P. 7-15, 2015.
- [6] C. Fracchia, P. Kogan, S. Amaro. Hornero: Aplicación para Gestión de Torneos de Programación, Multi-Paradigma, Multi-Plataforma, Multi-Lenguaje, Jornada. III Jornadas de Cloud Computing & Big DATA. RedUNCI, La Plata, 2015.
- [7] T. Perez, J. Plaza, C. Fracchia, A. Acuña. PowerPoint en el banquillo de los acusados. II Congreso Internacional de Investigación y Docencia de la Creatividad, Facultad de Ingeniería, UNCo, 2015.
- [8] C. Allan, S. Parra, A. Martins. Una Experiencia en la Enseñanza de la Matemática con Objetos de Aprendizaje. X Congreso sobre Tecnología en Educación & Educación en Tecnología. Corrientes, 2015.
- [9] Polsani, P., Use and abuse o reusable learnig journal of digital information, 2003. <https://journals.tdl.org/jodi/index.php/jodi/article/view/89/88>.
- [10] E. Zurbrigk, A. Alonso de Armiño, A. Martins, Accesibilidad Web y Educación para Todos, III Jornadas de educación mediada por Tecnología. Centro Universitario Zona Atlántica, UNCo, 2012.
- [11] Acceso a la Información Publica, Ley 26.653, <http://infoleg.gov.ar/infolegInternet/anexos/175000-179999/175694/norma.htm>
- [12] M. Pidd, Computer Simulation in Management Science, Wiley, 2004.

Metabuscador basado en agentes para grupos de estudiantes colaborativos

Carlos Pérez Crespo, María Martha Pérez Crespo y Rosanna Costaguta

Instituto de Investigación en Informática y Sistemas de Información (IISI)

Facultad de Ciencias Exactas y Tecnologías (FCEyT)

Universidad Nacional de Santiago del Estero (UNSE)

Avda. Belgrano (S) 1912 – Santiago del Estero CP 4200

charlyfpc@hotmail.com; mariamarthapc@yahoo.com.ar; rosanna@unse.edu.ar

Resumen

Muchas de las actividades que llevan a cabo grupos de estudiantes requieren que sus integrantes realicen individualmente búsquedas en la web para obtener información. A pesar de que muchos de los resultados obtenidos por cada uno de ellos se repiten, cada uno de esos resultados debe ser analizado para determinar su utilidad. Si se multiplica este proceso por la cantidad de individuos que conforman el grupo, se observa que puede producirse una considerable pérdida de tiempo y esfuerzo.

Dado el problema expuesto, se propone el desarrollo de un metabuscador basado en agentes, como herramienta web de búsqueda colaborativa que indica qué integrante analizó un determinado resultado, y que además permite asignar una valoración personal a cada resultado e incluir un comentario. Este metabuscador¹ podrá ordenar o rankear los resultados considerando una valoración individual asignada por cada integrante, y también por una valoración grupal obtenida promediando las valoraciones individuales.

¹ Los metabuscadores son motores de búsqueda que obtienen los resultados de otros múltiples motores de búsqueda. Permiten a los usuarios ingresar criterios de búsqueda una sola vez, y acceder a múltiples buscadores de forma simultánea [5].

El funcionamiento del metabuscador será validado mediante su uso por grupos de estudiantes colaborativos, en experiencias especialmente diseñadas para asignaturas de la carrera de Licenciatura en Sistemas de Información (FCEyT – UNSE).

Palabras clave: *Metabuscador, Agentes de Software, Búsqueda colaborativa, Grupos de estudiantes colaborativos.*

Contexto

En este artículo se presenta una de las líneas de investigación del proyecto 23/C097 “Sistemas de información web basados en agentes para promover el Aprendizaje Colaborativo Soportado por computadoras (ACSC)”, acreditado y financiado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Santiago del Estero (SECyT - UNSE) para el período 2012-2016. El proyecto es una continuación de la línea de investigación *Sistemas Adaptativos Inteligentes*, iniciada en 2005-2009 por el proyecto 23/C062 “Herramientas conceptuales, metodológicas y técnicas de la Informática Teórica y Aplicada”, continuada en 2009-2010 por el proyecto P09/C002 “Personalización en Sistemas de Enseñanza Virtual”, y en 2010-2011 por el proyecto 23/C089 “Fundamentos Conceptuales y Soportes Tecnológicos de la Informática Educativa”. Todos los proyectos citados fueron

acreditados y financiados por SECyT – UNSE.

Introducción

En estos últimos años, muchos trabajos se realizan a distancia por lo que Internet ha dejado de ser aquél lugar donde se consultaba información o se intercambiaban mensajes para pasar a ser un espacio abierto en el que todos pueden participar. Este concepto es la esencia de lo que se denomina Web 2.0, donde se concentran las conocidas herramientas 2.0 síncronas (Chat, Video llamadas, etc.) y asíncronas (Foros de Discusión, Wikis, Blog, etc.). Estas herramientas son soluciones tecnológicas que permiten a cualquier persona pasar de ser un mero receptor de información a ser partícipe de esa información, ya sea generándola, compartiéndola o mejorándola a través de redes de colaboración [3]. Estas nuevas capacidades de participación en la web posibilitó el surgimiento de los llamados Espacios Virtuales, sitios donde las herramientas 2.0 son utilizadas para aprendizaje y tutoría remota, generación de equipos acentuados por juegos, prácticas orientadas a escenarios, eventos globales multipropósito con complejas interacciones sociales, entre otros fines.

En particular en el ámbito educativo, los Espacios Virtuales posibilitan el trabajo de los estudiantes organizados en grupos, independizados de las variables tiempo y espacio. Para esto se valen de diversas herramientas web 2.0 disponibles para trabajar de forma eficaz colaborativamente y a distancia. Sin embargo, cuando los integrantes de un grupo deben realizar búsquedas, cada individuo dispara la suya de forma personal y sus resultados deben ser analizados individualmente para determinar su utilidad, luego realizar una puesta en común grupal, para finalmente tomar una decisión colectiva. Es indudable que muchos de los resultados obtenidos por para uno de los miembros pueden repetirse, y por ende, producirse considerables

pérdidas de tiempo y de esfuerzo en este proceso. Dado lo expuesto, en este trabajo se plantea desarrollar un metabuscador basado en agentes que dé apoyo a grupos de estudiantes colaborativos a fin de hacer más eficientes sus tareas de búsqueda individuales.

Como antecedentes de éste trabajo pueden citarse las siguientes investigaciones. En [1] y [4] se proponen metabuscadores para la recuperación de material web. De [1] resulta interesante el diseño e implementación de la arquitectura del metabuscador de párrafos propuesta, que recupera de la web documentos similares a fin de detectar plagios, y para la recuperación de ideas similares con altos niveles de precisión. De [4], rescatamos el uso de algoritmos de ranqueo para el ordenamiento y presentación del material web recuperado mediante la metabúsqueda. Otro trabajo relevante es [2], que propone un sistema multiagente para la búsqueda de productos en la web. Cabe destacar que no se han encontrado trabajos vinculados con el desarrollo de herramientas de búsqueda como apoyo a grupos colaborativos lo cual demuestra la originalidad de esta propuesta.

Líneas de Investigación, Desarrollo e innovación

Este trabajo se orienta en particular a la realización de búsquedas colaborativas de material digital en la web por parte de estudiantes que conforman un grupo. Se sabe que cada individuo tiene sus preferencias en cuanto a buscadores, técnica de búsqueda, etc., y que cuando efectúa una búsqueda generalmente usa un buscador al que le proporciona un conjunto de palabras clave y espera en respuesta una lista de resultados relacionados con esas palabras. Cuando la búsqueda debe realizarse en el marco de un grupo de estudiantes colaborativos, usualmente el proceder antes descrito se replica por parte de cada uno de sus integrantes. En esta situación, cada individuo debe revisar sus resultados, lo que

consume una considerable cantidad de tiempo y esfuerzo, para seguramente encontrar sólo unos cuantos resultados verdaderamente pertinentes. Además, dado que se trata de un grupo, los diferentes resultados obtenidos por la búsqueda efectuada por un integrante pueden repetirse en las búsquedas realizadas por otros integrantes del mismo grupo, independientemente del buscador utilizado, ya que dependerá altamente de las palabras clave ingresadas al disparar la búsqueda. Indudablemente, el esfuerzo, el tiempo y el resultado final del grupo se ven afectados cuanto mayor sea la cantidad de resultados repetidos. Sea de manera síncrona o asíncrona, una buena comunicación entre los participantes del grupo ayudaría a que esto no ocurra pero es normal que el tiempo empleado para indicar los resultados encontrados, pertinentes o no, por cada integrante al resto del grupo, aumente notoriamente cuando el número de miembros es mayor.

Por lo expuesto en los párrafos previos, la pregunta que guía esta investigación es la siguiente: ¿Cómo se puede hacer más eficiente la búsqueda de información en la web por parte de los integrantes de grupos de estudiantes colaborativo?

Resultados esperados y Objetivos

Este trabajo pretende, desde la perspectiva del proyecto de investigación en el que se halla inmerso, favorecer el desarrollo de conocimiento científico-tecnológico de relevancia para el desarrollo de sistemas de información web personalizados en el área del Aprendizaje Colaborativo Soportado por Computadora (ACSC), y también realizar propuestas metodológicas y desarrollos de módulos para sistemas de información web en ACSC (mediante la inclusión de modelos de usuario y agentes de software).

Desde la perspectiva puntual de este trabajo en sí mismo, se espera obtener un metabuscador basado en agentes, el cual pueda ser utilizado por grupos de

estudiantes que requieran realizar búsquedas colaborativas de material en la web. Los resultados puntuales esperados son:

- Estado del arte referido al trabajo en grupos de estudiantes colaborativos, búsquedas colaborativas, metabuscadores, agentes de software y sistemas de valoración de resultados.
- Algoritmo de ranqueo para determinar el orden de los resultados de búsqueda mediante una valoración grupal.
- Algoritmos para la consulta y actualización de los resultados de las búsquedas colaborativas almacenados en una Base de Datos.
- Algoritmos de funcionamiento de los agentes de software incluidos en el metabuscador.
- Herramienta de metabúsquedas basada en agentes que dé apoyo a las tareas de búsqueda de material web en grupos colaborativos.

Los objetivos de este trabajo quedaron enunciados de la siguiente forma:

- Propiciar búsquedas colaborativas eficientes.
- Favorecer la interacción entre los integrantes de grupos de estudiantes colaborativos.
- Generar una herramienta de búsqueda colaborativa de materiales web basada en agentes que permita ordenar los resultados considerando valoraciones individuales y grupales.

Para alcanzar estos objetivos, primero se realizará la búsqueda y análisis de bibliografía vinculada con la consideración trabajos de grupo, búsquedas colaborativas, metabuscadores, agentes de software y sistemas de valoración de resultados. Luego se llevará a cabo el diseño del metabuscador basado en agentes. Este metabuscador será implementado en un entorno de ACSC a fin de validar su funcionamiento mediante experiencias con estudiantes universitarios

reales. Actualmente, el trabajo está su etapa inicial de ejecución, es decir, realizándose el relevamiento bibliográfico correspondiente.

Por otra parte, resulta importante remarcar que la concreción de este proyecto está consolidando los esfuerzos y las investigaciones realizadas con anterioridad en el área del ACSC por parte de los integrantes de los proyectos citados en la sección Contexto.

Formación de Recursos Humanos

El equipo de trabajo de esta línea de investigación está compuesto por un docente investigador formado, y dos integrantes estudiantes. Estos estudiantes están desarrollando su tesis de grado para obtener el título de Licenciado en Sistemas de Información en el marco de esta investigación.

Referencias

- [1] Bravo Márquez, Felipe. 2010. *Diseño e Implementación de un Metabusador de párrafos para la recuperación de documentos similares en la web*. Tesis de grado. Universidad de Chile. [Http://repositorio.uchile.cl/handle/2250/103809](http://repositorio.uchile.cl/handle/2250/103809). [Último acceso: 15 de Marzo de 2016]
- [2] Espejo Bohórquez, Yeismer, Téllez, Magaly y Rodríguez, Jorge Enrique. 2012. Software basado en agentes inteligentes y servicios web para búsqueda de productos en la web. *Revista Tecnura*, Vol. 16 (31), pp. 114-125. [Http://www.scielo.org.co/pdf/tecn/v16n31/v16n31a11.pdf](http://www.scielo.org.co/pdf/tecn/v16n31/v16n31a11.pdf) [Último acceso: 15 de Marzo de 2016]
- [3] Observatorio Regional de la Sociedad de la Información de Castilla y León. 2012. *Herramientas para el Trabajo Colaborativo*. [Http://www.universoabierto.com/10824/herramientas-para-el-trabajo-colaborativo/](http://www.universoabierto.com/10824/herramientas-para-el-trabajo-colaborativo/) [Último acceso: 15 de Marzo de 2016]
- [4] Kuna, H., Rey, M., Martini, E., Solonezen, L., y Podkowa, L. 2014. Desarrollo de un Sistema de Recuperación de Información para Publicaciones Científicas del Área de Ciencias de la Computación. *Revista Latinoamericana de Ingeniería de Software*, Vol. 2(2), pp. 107-114. [Http://sistemas.unla.edu.ar/sistemas/redisla/ReLAIS/relais-v2-n2-107-114.pdf](http://sistemas.unla.edu.ar/sistemas/redisla/ReLAIS/relais-v2-n2-107-114.pdf). [Último acceso: 15 de Marzo de 2016]
- [5] López Curiel, Raúl. 2014. *Las TIC en el aula de Tecnología. Guía para su aplicación a la metodología de proyectos*. Serie Tecnología. Editorial Asociación para el Desarrollo del Profesorado.

Aplicaciones basadas en ontologías para asistir a la integración del aprendizaje móvil en escuelas de Catamarca

María Isabel Korzeniewski¹, Alejandro Sánchez², Germán Montejano^{2,3}

(1) Departamento de Informática, Facultad de tecnología y Ciencias Aplicadas.

Universidad Nacional de Catamarca

Maximio Victoria 55, Catamarca, Argentina. Tel.: +54 (383) 4435112

marisa.kor@gmail.com

(2) Universidad Nacional de San Luis

Ejército de los Andes 950 (5700), San Luis, Argentina. Tel.: +54-266-4520300 Int. 2128

fasanchez,gmonte}@unsl.edu.ar, web: <http://www.unsl.edu.ar>

(3) Universidad Nacional de La Pampa

Av. Uruguay 151, (6300) Santa Rosa, La Pampa, Argentina, Tel.: +54-2954-245220 Int. 7125

Resumen

El aprendizaje móvil introduce la posibilidad de aprovechar situaciones y contenidos fuera de las aulas, de manera de motivar e involucrar al alumno de nuevas formas. Sin embargo, su adopción es lenta o inexistente. Una de las barreras que explica esta lenta adopción, es la distancia que existe entre la conceptualización que maneja el docente, y la que se requiere para seleccionar una aplicación educativa móvil que se ajuste al entorno educativo de dicho docente.

Esta línea de I-D-I plantea utilizar aplicaciones basadas en ontologías para reducir esa distancia conceptual. Esto involucra desarrollar ontologías y prototipos de aplicaciones para anotar el software educativo, y para luego seleccionar los que respondan a una descripción.

Palabras clave: Aprendizaje móvil - Ontología - Aplicaciones móviles educativas - Dispositivos electrónicos personales

Contexto

Esta línea de investigación, desarrollo e innovación se desprende del Proyecto “Identificación y Clasificación de Herramientas TIC para la Gestión Estratégica” que se desarrolla dentro de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca, y del Proyecto “Ingeniería de Software” de la Facultad de Ciencias Físico-Matemáticas y Naturales de la Universidad Nacional de San Luis (UNSL)¹ y que incluye acciones de cooperación con otras universidades nacionales y del exterior en Europa y en América.

Introducción

El aprendizaje móvil puede caracterizarse como el aprendizaje a través de interacciones sociales y de contenido en múltiples contextos facilitadas por el uso de dispositivos

¹ <http://www.sel.unsl.edu.ar/pro/proyec/2012/index.html>

electrónicos personales (Crompton, 2013a). Este aprendizaje complementa el proceso en escuelas (y otros contextos de formación), introduciendo la posibilidad de aprovechar situaciones y contenidos fuera de las aulas. Permite incentivar el interés del alumno por medio de su participación activa, desarrollando competencias básicas, y por otra parte, potenciándose un enfoque constructivista (Romero, 2007; Valero, 2012; Aula Planeta, 2014). Es más, los alumnos pueden estar envueltos en el proceso de aprendizaje sin ser plenamente conscientes de ello (Burbules, 2012).

Sin embargo, la incorporación de este tipo de aprendizaje en la escuela es pobre o inexistente. Las principales barreras incluyen dificultades en el soporte financiero, temor a violación de datos privados, y estilos de educación (UNESCO, 2012; Crompton, 2013b). Nos enfocamos en la reticencia a incorporar elementos que puedan ser percibidos como disruptivos en el proceso, y más allá de ella, existe conocimiento y habilidades a los que los docentes deben acceder para poder integrar objetos educativos adecuados del aprendizaje móvil al proceso de enseñanza-aprendizaje (Khe Foon, 2007). De hecho, parece existir una relación entre estas dos barreras; hay resultados que indican que la reticencia a integrar tecnología se origina por la falta de conocimiento y habilidades adecuadas (Hughes, 2005).

El enfoque clásico, para capacitar a los docentes, es el dictado de cursos en tecnologías y en metodologías pedagógicas que exploten su utilización.

En nuestra línea de investigación, desarrollo e innovación (I-D-I), argumentamos que este enfoque es insuficiente para lograr la integración buscada. Existe un número de parámetros que se mantienen ideales durante un curso, y que en entornos reales crecen en

complejidad y se convierten en una nueva barrera que el docente debe enfrentar para llevar al aula el aprendizaje móvil.

Por un lado, existe un número de parámetros tecnológicos que se deben considerar. Entre estos mencionamos las capacidades de los dispositivos electrónicos personales de los alumnos (tipos y tamaños de pantallas, tipos de conectividad, capacidades de procesamiento, multimedia y de geolocalización, etc.), el ritmo vertiginoso de la evolución de las tecnologías, la infraestructura de comunicación disponible en la región (un problema registrado en Catamarca es la falta o intermitencia en la conectividad en centros educativos).

Por el otro lado, existe un gran dinamismo en las metodologías pedagógicas relacionadas al aprendizaje móvil. El dinamismo de las tecnologías y las posibilidades de interacción que permiten, llevan a la aparición de nuevas estrategias pedagógicas y las correspondientes metodologías.

A su vez, las aplicaciones móviles educativas varían dependiendo (entre otros parámetros) de los requerimientos tecnológicos y de las metodologías pedagógicas subyacentes.

La capacitación del docente no termina con los cursos, sino que comienza allí. El docente debe poder acceder a contenidos relacionados a los problemas específicos de su contexto, para poder continuar así su capacitación en estos, hasta lograr integrar el aprendizaje móvil a su entorno educativo.

Sin embargo, existe una distancia conceptual y lingüística que dificulta esta tarea. La rápida evolución y la multiplicidad de instituciones, industrias y comunidades envueltas, lleva a la existencia de una gran cantidad de taxonomías pobremente relacionadas. La falta de una vista global dificulta la tarea

del docente. Este puede desconocer la taxonomía con la cual se ha clasificado una aplicación educativa móvil (por ej., de acuerdo a sus parámetros tecnológicos o metodológicos), o estar familiarizado con otra taxonomía con la cual haya una superposición que desconoce.

A su vez, las descripciones disponibles de las aplicaciones educativas móviles están expresadas en lenguaje natural. De esta forma, no solo puede estar utilizando taxonomías desconocidas para el docente, sino que un lenguaje sesgado por, por ejemplo, tendencias de mercadeo o modismos locales.

Los motores de búsqueda clásicos prestan limitado soporte frente a este problema. El docente puede recurrir a los tradicionales buscadores de internet, pero estos normalmente retornarán un número inmanejable de resultados irrelevantes (Aguilar, López, Sosa, 2009).

Numerosos trabajos plantean el uso de ontologías (Gruber, 1993) para anotar y recuperar recursos (Hertel, 2009). El objetivo es utilizar un razonador para retornar resultados que satisfagan una descripción conceptual del ítem buscado. Los resultados obtenidos de esta forma serán los clasificados dentro de la ontología subyacente como instancias de dicha descripción conceptual. Muchos de estos trabajos, como ser Aguilar et al. (2009), o Romero (2007) o Martín (2009), utilizan a su vez lenguajes estándares de especificación de ontologías basados en lógicas descriptivas (Baader, 2003).

Estas ontologías que permiten describir el recurso buscado son generalmente específicas para un dominio, y frecuentemente utilizan conceptos definidos en ontologías de alto nivel (Guarino, 1998). Las ontologías de dominio a su vez son utilizadas cuando se desarrollan ontologías para soportar una aplicación particular. Las relaciones entre

ontologías se establecen a través de técnicas llamadas de “ontology matching” en lenguajes específicos de correspondencia (Euzenat, 2013).

Líneas de Investigación, Desarrollo e Innovación

Esta línea plantea investigar, desarrollar e innovar sobre aplicaciones basadas en ontologías para asistir al docente en la selección de aplicaciones educativas móviles de acuerdo a los parámetros tecnológicos disponibles, y a las metodologías pedagógicas que mejor se ajusten al entorno educativo de Argentina y Catamarca, para promover el aprendizaje centrado en el alumno a través de un uso distinto de una tecnología que maneja diariamente. Los principales ejes de trabajo son:

- La participación de expertos del dominio de la región como usuarios y fuente de conocimiento.
- La selección y ajuste de una metodología de desarrollo de ontologías, y software asociado.
- La utilización de herramientas para el desarrollo de ontologías especificadas en lenguajes basados en lógicas descriptivas.

Resultados y Objetivos

Esta línea de I-D-I se encuentra en una etapa inicial. Como objetivo se plantea asistir al docente en la selección de aplicaciones educativas móviles. Para ello es necesario investigar y desarrollar prototipos de:

- Ontologías de dominio y/o aplicación para dispositivos electrónicos personalizados, metodologías pedagógicas relevantes, y para el aprendizaje móvil, ajustadas al

contexto educativo de la Argentina y la provincia de Catamarca.

- Aplicaciones que permitan a docentes más experimentados en el aprendizaje móvil, describir aplicaciones educativas móviles utilizando las ontologías desarrolladas, sin ser expertos en ontologías.
- Aplicaciones que permitan la selección de aplicaciones educativas móviles en base a la descripción de las necesidades del entorno educativo, especificada a partir de las ontologías provistas, por un docente sin experiencia en ontologías.

Formación de Recursos Humanos

El equipo de proyecto está formado por: dos Mg en Ingeniería de Software (2015) y que actualmente están comenzando su doctorado, una Especialista en TIC en Educación, dos Licenciadas en Sistemas de la Información, que han terminado de cursar la Maestría en Ingeniería de Software y están desarrollando sus tesis, dos alumnos de la carrera Ingeniería en Informática, que están desarrollando sus tesis de grado. Es importante destacar el apoyo académico y científico para los temas desarrollados en la formación de recursos humanos calificados ya que son carreras de posgrado categorizadas "A" para los desarrollos de las tesis de posgrado y carreras de grado acreditadas por 6 años para los desarrollos de las tesis.

Referencias

Aguilar-Lopez, Dulce; Lopez-Arevalo, Ivan; Sosa-Sosa, Víctor (2009, enero-febrero). *Uso de ontologías para la mejora de resultados de motores de búsqueda web*. El profesional de la

información, vol. 18, núm. 1, pp. 34-40.

Aula Planeta (2014) *Perspectiva 2014: Tecnología y Pedagogía en el aula*. Editorial Planeta. S.A.U. disponible en: http://biblioteca.ucv.cl/site/colecciones/manuales_u/aulaPlaneta_Perspectivas-2014.pdf

Baader, Franz; Diego, Calvanese; McGuinness, Deborah; Nardo, Daniele, and Peter F. Patel-Schneider (Eds.). (2003). *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, New York, NY, USA.

Burbules N. (2012) *Aprendizaje Ubicuo*, entrevista realizada por IIPPEE - UNESCO, Buenos Aires. Argentina Video disponible en: <http://www.iipe-buenosaires.org.ar/node/64>.

Crompton, H. (2013a). *Mobile learning: New approach, new theory*. In Z. L. Berge & L. Y. Muilenburg (Eds.), *Handbook of mobile learning* (pp. 47-57). Florence, KY: Routledge. (Awarded the 2014 Association for Educational Communications and Technology (AECT) Division of Distance Learning (DDL) Distance Education Book Award)

Crompton, H. (2013b). *Research windows: The benefits and challenges of mobile learning*. *ISTE Learning and Leading with Technology*. 41 (2), 38-39

Gruber, Thomas R. (1993). *A Translation Approach to Portable Ontology Specifications*. En: *Knowl. Acquis.* 5.2, págs. 199-220. ISSN : 1042-8143.

- Guarino, N.** (1998). *Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy*. 1st. Amsterdam, The Netherlands, The Netherlands: IOS Press. ISBN : 9051993994.
- Hertel, A.; Jeen Broekstra; Heiner Stuckenschmidt** (2009). *RDF Storage and Retrieval Systems. Handbook on Ontologies*. Part of the series International Handbooks on Information Systems pp 489-508.
- Hughes, J.** (2005). *The Role of Teacher Knowledge and Learning Experiences in Forming Technology-Integrated Pedagogy*. Journal of Technology and Teacher Education. 13 (2), pp. 277-302. Norfolk, VA: Society for Information Technology & Teacher Education.
- Jérôme, Euzenat; Pavel Shvaiko.** (2013). *Ontology Matching* (2nd ed.). Springer Publishing Company, Incorporated
- Khe Foon Hew, T Brush** (2007). *Integrating technology into K-12 teaching and learning: Current knowledge gaps and recommendations for future research*. Educational Technology Research and Development 55 (3), 223-252
- Martin, A. et al.** (2009) *Perfil inteligente de ontologías para la recuperación de la información*. Norma Ponencia presentada en, IX Congreso Isko-España, en Valencia. España disponible en: <https://dialnet.unirioja.es/descarga/articulo/2924493.pdf>
- Romero Llop, R.** (2007) *Especificación Owl De Una Ontología Para Teleeducación En La Web Semántica* Tesis para la obtención del grado de: Doctorado, Universidad Politécnica de Valencia
- Scopeo** (2011-Noviembre). *M-learning en España, Portugal y América Latina*, Monográfico SCOPEO, no 3. ISSN 1989-8266. Universidad de Salamanca. Centro Internacional de Tecnologías Avanzadas, Observatorio de la Formación en Red. en: <http://scopeo.usal.es/investigacion/monograficos/scopeom003>
- UNESCO** (2012). *Aprendizaje móvil para docentes en América Latina: Análisis del potencial de las tecnologías móviles para apoyar a los docentes y mejorar sus prácticas*. ISSN 2305-8617.
- Valero, CC.** et al. (2012) *Tendencias actuales en el uso de dispositivos móviles en educación* Educational Portal of the Americas – Department of Human Development, Education and Culture© OEA-OAS ISSN 0013-1059 La Educ@ción Digital Magazine N 147 – www.educoas.org. Disponible en http://educoas.org/portal/la_educacion_digital/147/pdf/ART_UNNED_EN.pdf.

Ontologías para entornos personalizados de aprendizaje ajustados al contexto educativo argentino

Ana del Prado¹, Alejandro Sánchez², Germán Montejano^{2,3}

¹Facultad de Tecnología y Ciencias Aplicadas – Universidad Nacional de Catamarca
Maximio Victoria 55 - San Fernando del Valle de Catamarca - TEL. +54 383 4435112
anadelprado@tecno.unca.edu.ar

²Universidad Nacional de San Luis
Ejercito de Los Andes 950, San Luis, Argentina - Tel: +54 266 4520300
{asanchez,gmonte}@unsl.edu.ar

³Universidad Nacional de la Pampa
Av. Uruguay 151, (6300) Santa Rosa, La Pampa, Argentina, Tel.: +54-2954-245220 Int. 7125

Resumen

Un entorno personalizado de aprendizaje (EPA) se centra en las preferencias del alumno, sus características, sus conocimientos previos, y su comportamiento, y le ofrece, a su vez, actividades alternativas para fortalecer contenidos en los que se muestra débil. Normalmente se utiliza como complemento a procesos de enseñanza-aprendizaje escolares, permitiendo motivar el avance de alumnos con conocimientos tanto por encima, como por debajo, del promedio.

Esta línea plantea la investigación y desarrollo de ontologías sobre el conocimiento necesario para responder a las preguntas formuladas durante el funcionamiento de un EPA en el contexto de la educación argentina.

Palabras clave: ontología, EPA, TIC.

Contexto

Esta línea de investigación, desarrollo e innovación (I-D-I) se desprende del proyecto “Exploración, identificación, selección de herramientas TIC para la gestión estratégica”, de la Universidad

Nacional de Catamarca – Secretaria de Ciencia y Tecnología y del proyecto “Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la Profesión del Ingeniero de Software” de la facultad de Ciencias Físico – Matemáticas y Naturales de la Universidad Nacional de San Luis y que incluye acciones de cooperación con otras universidades nacionales y del exterior en Europa y en América.

Introducción

Una de las principales necesidades de los docentes es provocar aprendizajes que perduren y donde el alumno se encuentre motivado. Sin embargo, las limitaciones de tiempo y espacio en las aulas dificulta esta posibilidad. Los alumnos con más conocimientos en un determinado tema se deben adecuar al nivel de aprendizaje del grupo, desmotivándose y sin poder avanzar en su aprendizaje.

Relacionado a esto se encuentra la dicotomía entre las necesidades y los medios, es decir, no existe un acompañamiento entre el desarrollo tecnológico y la didáctica.

Por ello, surge la necesidad de generar un entorno que posea las herramientas

necesarias y adecuadas para que el alumno pueda aprender por sí mismo en forma eficiente (Fainholc, 1999). El desafío consiste en generar entornos no convencionales, donde el alumno pueda participar, crezca su capacidad de pensamiento, y ejercite su práctica.

El concepto de EPA emerge para atender esta necesidad (Adell, 2010; Wilson, 1995). El objetivo es aprovechar las TIC para ofrecer al alumno un entorno centrado en sus preferencias, sus características, y su comportamiento, ofreciéndole a su vez alternativas para fortalecer los temas en que se muestra débil. De esta forma se evitan, o complementan, los procesos estándares dirigidos a todos los alumnos, con los mismos contenidos, actividades, estímulos y evaluaciones.

Entre los aspectos fundamentales que un EPA debe considerar se encuentran el nivel de conocimiento del alumno, sus preferencias de estudio, las estrategias de aprendizaje adecuadas para él, y el objetivo primordial de motivar su avance. No se limita a un aprendizaje cerrado, o lineal, de los contenidos, como ocurre tradicionalmente, lo cual se traduce en una mejora en la calidad del aprendizaje.

La motivación del alumno se puede lograr mediante actividades que sean adecuadas a su nivel de conocimiento y preferencias de estudio, sin depender de un lugar o tiempo fijado.

El EPA debe establecer un plan personalizado de aprendizaje, donde los estudiantes trabajen de manera autónoma, dominando contenidos y habilidades concretas que le permitan actuar en diferentes situaciones, resolviendo los problemas que se le presentan.

Por lo tanto, los conocimientos de los EPAs, deben poder adaptarse a las necesidades educativas argentinas y responder a las preguntas que siguen.

¿Cuál es el nivel de conocimiento del alumno sobre un tema en particular?

¿Qué actividades sugerir a un alumno en base a sus preferencias y nivel de conocimiento?

¿Qué estrategias de aprendizaje le resultan adecuadas al alumno?

¿Cómo motivar al alumno que se esfuerza por adquirir nuevos conocimientos?

La implementación de un EPA debe permitir responder a las preguntas que emergen al perseguir el objetivo de motivar al alumno.

Las ontologías (Borst & Akkermans, 1997; Gruber 1993) se han utilizado como enfoque para proveer soluciones a este tipo de problema (Albano y colegas 2007, Albano y colegas 2015). Una ontología es una representación formal y compartida que brinda una alternativa para atender a problemas como el enunciado. Resulta más flexible, ya que es una representación explícita de las teorías subyacentes que puede ser modificada para responder a cambios en las mismas, seguramente con mayor facilidad que el código fuente de una aplicación.

Estos enfoques, en general, implican ontologías de distintos tipos (Guarino, 1998). Se suele requerir el desarrollo/reutilización de ontologías de dominio, y el desarrollo de ontologías de aplicación. En ambos casos, las ontologías pueden incorporar entidades de ontologías de alto nivel. Las relaciones entre estas ontologías se desarrollan conforme a la disciplina de correspondencia entre ontologías (Euzenat & Shvaiko, 2013).

A su vez, existen metodologías de desarrollo que permiten la participación, supervisión, y evaluación de los productos por parte de expertos del dominio. Lo cual es particularmente importante de cara al objetivo de que los

resultados se ajusten al entorno educativo argentino. Entre dichas metodologías podemos citar (Grüninger & Fox, 1995), (Uschold & King, 1996), Kactus (Bernaras y colegas, 1996), Methontology (Gómez-Pérez y colegas, 1996), On-to-Knowledge (Staab y colegas, 2001), Terminae (Corcho y colegas, 2003) y Refseno (Tautz & Wangenheim, 1998).

Resulta importante resaltar el creciente interés en el desarrollo y uso de ontologías en el dominio E-Learning en el periodo 2008 a 2012 como muestra (Al-Yahya y colegas, 2015), el volumen de trabajos publicados ha crecido en google scholar comparado con las bibliotecas digitales de ACM, IEEE y Science Direct. Estas tendencias indican un creciente interés respecto al uso de ontologías en los sistemas de e-learning, lo cual destaca la relevancia que adquiere esta temática en los últimos tiempos.

Se han identificado esfuerzos desarrollando ontologías para e-learning. Entre los trabajos más relevantes a nivel mundial podemos mencionar los que siguen.

En (Prमितasari y colegas, 2009) se desarrolla una ontología que comprende los datos personales del estudiante, estilo de aprendizaje y desempeño de los estudiantes. Estos aspectos se consideran importantes para describir el modelo de los estudiantes pero no son aplicados a ningún caso en particular, sino que pretende ser un complemento de otras ontologías de aprendizaje.

La referencia (Paneva, 2006) presenta una ontología que modela al estudiante, comprende la información del mismo, y el estilo de aprendizaje de los estudiantes en el dominio de aprendizaje. Se muestra como las soluciones mediante web semántica pueden completar el desarrollo de una aplicación de aprendizaje.

En (Šimún y colegas, 2007) se propone modelo de dominio y usuario de un

sistema adaptativo basado en la web. Utiliza los principios de la web semántica para los modelos de representación y manipulación. La personalización se basa en las características del usuario que se almacenan en el modelo de usuario. Hay tres tipos de atributos de los estudiantes: preferencias, intereses del estudiante y la experiencia previa, o conocimiento de los estudiantes.

Entre los esfuerzos relacionados al tema en el contexto americano, se encuentran los que siguen.

En (Garnica, y colegas, 2014) se identifican las necesidades del proceso de aprendizaje, para crear sistemáticamente cada una de las actividades, se obtiene un modelo ontológico a través de Methontology y se implementa con la herramienta Protégé-OWL. Esto permite generar una base de conocimientos para diseño de Secuencias Didácticas.

En (García y colegas, 2014) se utiliza Methontology para el modelo de aprendizaje colaborativo y los elementos que lo conforman, así como las reglas de los procesos de gestión en la interacción según las pautas propias de un ambiente empresarial.

En Argentina, encontramos el trabajo (González, & Durán, 2014), donde se consideran las características personales y el contexto del alumno, se tiene en cuenta datos personales del alumno, conocimiento actual y el rendimiento en el curso, que es una clase que no está definida en el trabajo.

En resumen, un análisis preliminar de la literatura indica que en las aplicaciones educativas, la actualización o adaptación se realiza sólo sobre la base de los conocimientos del alumno y la serie de trabajos analizados no presentan descripciones con alto nivel de detalle de clase y propiedad de ontologías utilizadas. Por otra parte, las ontologías

no son accesibles en línea con el propósito de estudios detallados.

Líneas de investigación, desarrollo e innovación

Esta línea de I-D-I plantea investigar y desarrollar ontologías para el dominio de la personalización del aprendizaje, y ontologías de aplicación, refinamiento de la primera, para atender los requerimientos de un EPA específico hipotético a ser desarrollado.

Esta línea contrasta con la posibilidad de dejar este conocimiento implícito en el código fuente de una implementación particular de EPA. Dicho conocimiento incluye la modelización de los aspectos fundamentales de la personalización del aprendizaje y las teorías para responder a las preguntas planteadas. Esta implementación la realizaría un programador, posiblemente no experto en el dominio, y no sería de fácil reutilización en otras aplicaciones. A su vez, los eventuales cambios que surgen a partir de posibles mejoras en la didáctica y teorías pedagógicas, requieren un mantenimiento que se vería dificultado con esta representación implícita. Es más, en el ámbito de la enseñanza-aprendizaje, es difícil tener en cuenta todas las variables, en educación hay contingencias impredecibles, que dependen, por ejemplo, del avance tecnológico, de la infraestructura disponible, de los estudiantes y de los educadores.

Objetivos y resultados esperados

Esta línea de I-D-I se encuentra en su etapa inicial. Su objetivo es incorporar, reutilizar y adaptar los conocimientos de los EPAs conforme a las necesidades educativas locales para el nivel de conocimiento de un alumno, sus

preferencias, y las estrategias de aprendizaje adecuadas a él, de manera de emplear EPAs, donde el alumno pueda adquirir conocimientos por niveles y acorde a sus preferencias con actividades según estrategias de aprendizaje y que lo motiven a avanzar en el proceso.

Las ontologías buscarán especificar el vocabulario, las relaciones, y las teorías generales del problema y soluciones al mismo. Permitirán inferir respuestas según determinados parámetros, y a un conjunto de axiomas que hacen explícita la teoría detrás de la personalización del aprendizaje. Las respuestas se materializarán, por ejemplo, como sugerencias de actividades, tareas, o experiencias que apunten a motivar a un alumno en particular. Los resultados esperados incluyen:

- Ontologías sobre EPA para el contexto educativo argentino.
- Aplicación prototípica que resulte una respuesta innovadora a los desafíos de calidad educativa de nuestro país.

Formación de Recursos Humanos

Esta línea de I-D-I continúa con la formación en investigación, de los integrantes del proyecto de la Facultad de Tecnología y Cs. Aplicadas de la UNCa. Está previsto que dos integrantes del grupo desarrollen en el tema, su Tesis de Maestría en Ingeniería de Software, así como también las tesinas de dos alumnos de grado. Es importante destacar el apoyo académico y científico para los temas desarrollados en la formación de recursos humanos calificados, ya que son carreras de posgrado categorizadas "A" para los desarrollos de las tesis de posgrado y carreras de grado acreditadas por 6 años para los desarrollos de las tesinas.

Referencias

- Adell, J., & Castañeda, L. (2010). Los Entornos Personales de Aprendizaje: una nueva manera de entender el aprendizaje.
- Albano, G., Gaeta, M., & Ritrovato, P. (2007). IWT: an innovative solution for AGS e-learning model. *Int. Journal of Knowledge and Learning*, 209-224.
- Albano, G., Miranda, S., & Pierri, A. (2015). Personalized Learning in Mathematics. *Journal of e-Learning and Knowledge Society*, 11(1).
- Al-Yahya, M., George, R., & Alfaries, A. (2015). Ontologies in E-learning. *Int. J. Softw. Eng. Appl*, 9(2), 67-84.
- Bernaras, A., Laresgoiti, I., & Corera, J. (1996). Building and Reusing Ontologies for Electrical Network Applications'. pp. 298-302.
- Borst, P., & Akkermans, H. (1997). An ontology approach to product disassembly. In *Knowledge Acquisition, Modeling and Management* pp. 33-48.
- Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point?. *Data & knowledge engineering*, 41-64.
- Euzenat, J., & Shvaiko, P. (2013). Classifications of ontology matching techniques. In *Ontology matching* pp. 73-84. Springer Berlin Heidelberg.
- Fainholc, B. (1999). La interactividad en la educación a distancia. *Paidós*.
- García, A. C. M., Saldivia, B. S., & Monzón, G. P. (2014). Un modelo ontológico para el aprendizaje colaborativo en la educación interactiva a distancia. *Educere*, 18(61), 449-460.
- Garnica, C. C., Sierra, E. A., Martínez, B. B., Márquez, P. C., & Cruz, J. L. G. Elaboración de una ontología para apoyar el diseño de secuencias didácticas basadas en competencias en la práctica del docente de educación media superior. *Avances en la Ingeniería del Lenguaje y del Conocimiento*, 115.
- Gómez-Pérez, A., Fernández, M., & Vicente, A. D. (1996). Towards a method to conceptualize domain ontologies.
- González, G., & Durán, E. B. (2014). Modelo del estudiante para sistemas de aprendizaje ubicuo: representación por medio de ontologías. IX Cong. TE & ET.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), 199-220.
- Grüninger, M., & Fox, M. S. (1995). *Methodology for the Design and Evaluation of Ontologies*.
- Guarino, N. (1998). Formal ontology in information systems: Proceedings of the first international conference, 6-8.
- Paneva, D. (2006). *Ontology-Based Student Modelling*.
- Pramitasari, L., Hidayanto, A. N., Aminah, S., Krisnadhi, A. A., & Ramadhanie, A. M. (2009). Development of student model ontology for personalization in an elearning system based on semantic web. In *International Conference on Advanced Computer Science and Information Systems*, 7-8.
- Šimún, M., Andrejko, A., & Bieliková, M. (2007). Ontology-based models for personalized e-learning environment. 5th Int. Conf. on Emerging elearning Technologies and Applications.
- Staab, S., Studer, R., Schnurr, H. P., & Sure, Y. (2001). Knowledge processes and ontologies. *IEEE*, (1), 26-34.
- Tautz, C., Wangenheim, G., & Refseno, C. (1998). A representation formalism for software engineering ontologies. *Fraunhofer IESE*, 15, 1-151.
- Uschold, M., & King, M. (1995). Towards a methodology for building ontologies pp. 15-30. Edinburgh.
- Wilson, B. G. (1995). Metaphors for instruction: Why we talk about learning environments. *Educational Technology-Saddle Brook NJ-*, 35, 25-25.

Red AUTI Red de Aplicaciones y Usabilidad de la Televisión Digital Interactiva (TVDi)

María José Abásolo¹², Armando De Giusti¹, Cecilia Sanz¹, Patricia Pesado¹,
Sabrina Martorelli¹³, Verónica Artola^{1,4}, Marcelo Naiouf¹, Alejandra
Zangara¹, Graciela Santos⁵, Sandra Casas⁶

¹Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata (UNLP)

{mjabasolo, degiusti, csanz, ppesado, smartorelli, vartola, mnaiouf}@lidi.info.unlp.edu.ar,
alejandra.zangara@gmail.com

²Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

³Becaria Tipo A UNLP

⁴Becaria Doctoral CONICET

⁵ Núcleo de Investigación en Ciencias con Tecnología. Departamento de
Formación Docente, Facultad de Ciencias Exactas
Universidad Nacional del Centro de la Pcia. de Bs. As (UNICEN)
nsantos@exa.unicen.edu.ar

⁶Universidad Nacional de la Patagonia Austral (UNPA)
scasas@unpa.edu.ar

Resumen

Desde el año 2012 diversas universidades iberoamericanas conforman la Red de Aplicaciones y Usabilidad de la Televisión Digital Interactiva (RedAUTI) de la CYTED coordinada por la UNLP. La línea de investigación y desarrollo presentada consiste en estudiar, desarrollar y evaluar aplicaciones de Televisión Digital Interactiva (TVDi). Uno de los principales objetivos es la formación de recursos humanos y fortalecimiento de la investigación mediante el trabajo intergrupar entre diferentes instituciones nacionales y extranjeras.

Palabras Clave: Televisión digital, TVDi, Interactividad

Contexto

La RedAUTI¹ cuenta con la participación de 225 investigadores de 36 grupos de investigación (29 universidades y 7 empresas) de España, Portugal y 10 países latinoamericanos (Argentina, Brasil, Chile, Colombia, Cuba, Ecuador, Guatemala, Perú, Uruguay, Venezuela). La RedAUTI fue creada y financiada por el Proyecto 512RT0461 Red de Aplicaciones y Usabilidad de la Televisión Digital Interactiva (RedAUTI) [1] del Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED) durante el período 2012-2015. Recientemente fue adjudicado un proyecto de la convocatoria Redes IX de la Secretaría de Políticas

¹ <http://redauti.net>

Universitarias del Ministerio de Educación Argentino en el que se continuará profundizando en esta línea de investigación.

Introducción

La televisión digital abre la oportunidad de crear aplicaciones interactivas, de ahí la denominación de Televisión Digital Interactiva (TVDi). La interactividad permite un diálogo donde el participante tiene la capacidad de intervenir en los programas o servicios que recibe. Con la TVDi el consumidor puede pasar de ser un espectador pasivo a convertirse en un participante activo. A través de la TVDi puede ser posible acceder a un conjunto de servicios que abarcan diversos campos como comercio, gestión administrativa, entretenimiento y educación. Esto supone una alternativa al uso de la computadora e internet, lo cual puede facilitar el acceso a la sociedad de la información en sectores que aún no disponen de esa tecnología. La interactividad con canal de retorno permite no sólo ver contenidos adicionales a la programación y navegar por ellos, sino también enviar respuestas por parte de los usuarios, e incluso comunicarse con otros usuarios. Por ejemplo, permite a los usuarios participar en concursos, votar, realizar solicitudes o enviar mensajes a través de la aplicación interactiva. Si bien es una herramienta que sin dudas puede atraer tanto a las políticas públicas como a las políticas educativas, también puede ser de interés al sector comercial que puede conectar con sectores específicos de las audiencias y ofrecer servicios comerciales particularizados.

En el año 2009 se creó el Sistema Argentino de Televisión Digital Terrestre, y el Consejo Asesor de la Televisión Digital Terrestre con el objetivo de guiar el proceso de migración tecnológica hacia la

televisión digital en la República Argentina hasta el apagón analógico o encendido digital previsto en el 2019. Durante los años que lleva adelante la implementación de la televisión digital en Argentina, la política de distribución de licencias genera tensiones entre el desarrollo del plan nacional de servicios de comunicación audiovisual digitales y en la administración del espectro radioeléctrico [Gon13][Gon14][Gon15].

En Argentina los primeros abordajes en materia de interactividad llegaron desde la investigación académica en universidades nacionales y en experiencias pedagógicas y de transferencia tecnológica. En los casos desarrollados, la noción de interactividad se relaciona con la navegación, la presentación de contenidos en formatos multimedia (gráficos, sonoros, audiovisuales y transferencia de datos). Sin embargo, y si bien se han desarrollado algunas pruebas, falta desarrollar de manera concreta la articulación del *set-top-box* con la conectividad a internet, establecer un canal de retorno y generar una interactividad plena donde el televidente pueda intervenir en la producción de los contenidos y sus intervenciones sean consideradas, tal como ha sucedido en otros países del mundo. A las experiencias que se desarrollan actualmente en nuestro país se las denomina de interactividad local, donde el usuario puede acceder a contenidos interactivos pero no puede enviar datos de vuelta. Los ejemplos más comunes de aplicaciones interactivas locales son las guías electrónicas de programación, el teletexto digital, o la información adicional.

Artola et al. [Art15] presentan una recopilación de aplicaciones de TVDi que se están desarrollando en Argentina y se las clasificó de acuerdo a una serie de criterios específicos tales como: si están orientadas a ofrecer servicios, a la

educación, o a la salud, el tipo de interactividad y el contexto en el que fue desarrollada. Esto da cuenta de la intervención de las universidades argentinas en esta temática.

El sector productivo aún no dejó ver sus avances. El desarrollo de herramientas que permitan la generación automática o semi-automática de las aplicaciones interactivas [Cas14][Cas15][Mir16] contribuirá a que la industria conformada principalmente por pequeños y medianos productores de contenidos, se inicien en la generación de contenidos interactivos, de manera fácil y a muy bajo costo. Las herramientas son esencialmente generadores de código, que a su vez aplican diversas técnicas de ingeniería de software: patrones de diseño de interactividad, líneas de productos de software, métodos ágiles, etc.

El término t-learning se refiere al aprendizaje interactivo a través de un televisor, es decir aplicaciones interactivas para TVDi enfocadas principalmente a apoyar los procesos de enseñanza y aprendizaje. El uso de la TVDi en educación se encuentra aún en pleno desarrollo e investigación, en particular en los países latinoamericanos. Zangara et al. [Zan13] describen la diferenciación entre interacción e interactividad, y una vinculación entre interactividad educativa y el concepto de distancia transaccional. En [Zan14] se presentan algunas perspectivas que podrían tenerse en cuenta en la producción de materiales educativos en el campo de la televisión digital interactiva, analizando componentes propios de la teoría de distancia transaccional de la educación a distancia como son estructura, diálogo y autonomía del estudiante y, dentro de estos componentes, la interactividad, personalización y asistencia al estudiante.

Se están desarrollando algunos ejemplos de educación a distancia como el caso presentado por Martorelli et al. [Mar16] “HistoTV”, una aplicación educativa para TVDi sobre Histología Animal, a partir de la cual es posible acceder a preparados microscópicos de tejidos animales de manera tal de ayudar a los alumnos en la identificación de estos tejidos, a partir de las observación de diferentes muestras. La aplicación está compuesta por videos que presentan los distintos tejidos animales. Cada video es una visualización en un microscopio virtual de preparados histológicos de tejidos acompañados de explicaciones del docente de la materia. Además se incluyen una serie de preguntas de autoevaluación, a manera de juego, donde se le pide al alumno identificar los tejidos animales en diferentes imágenes que se le presentan.

La evaluación de este tipo de aplicaciones educativas incluye el conocimiento de sus posibilidades y limitaciones así como los esquemas de uso previstos en su diseño y aquellos otros que los usuarios emplean [San13]. En esta investigación también es de interés analizar qué es lo que los usuarios esperarían poder hacer con la aplicación, a efectos de maximizar las potencialidades interactivas de contenidos para TDVi. Se ha definido una metodología para estudiar la usabilidad desde la perspectiva de las interacciones que ejecuta la persona y las que quisiera hacer, reflejadas en aquello que dicen mientras interactúa [San16]. Considera la reconstrucción de las narrativas de interacción a partir de las acciones en pantalla, y se han identificado algunas perturbaciones que promueven diferentes secuencias de navegación. Además de evaluar las actuaciones de los usuarios, se propone estudiar la relación coste/beneficio o del Retorno de la Inversión (ROI) para distintos grados de interactividad.

Líneas de Investigación, Desarrollo e Innovación

La RedAUTI se propone el intercambio de aplicaciones interactivas desarrolladas para su evaluación en los diferentes contextos. El objetivo general es apoyar la formación de los recursos humanos, estimular la investigación de forma transdisciplinaria y fortalecer los grupos de investigación de las Universidades iberoamericanas en temas relacionados con TVDi, avanzando en el diseño, implementación y despliegue de aplicaciones, servicios y producción de contenidos para TVDi, en sus múltiples plataformas, de código abierto, de interés colectivo, para dar solución a problemas del contexto iberoamericano.

Las líneas de investigación y desarrollo principales son:

- Contenidos y Aplicaciones Interactivas para TVDi, con especial interés en T-learning
- Relación Interactividad y Modelos pedagógicos de T-learning
- Herramientas informáticas para el desarrollo de aplicaciones para TVDi
- Experiencias de usuario con la TVDi: usabilidad y accesibilidad

Resultados y Objetivos

Los objetivos específicos de la RedAUTI son:

1- Identificar problemas del contexto iberoamericano que pueden ser solucionados con la construcción de contenidos, aplicativos y servicios soportados en las tecnologías de la TVDi. En particular, ofrecer soluciones de código abierto, de interés colectivo, asociadas a la mejora del bienestar de la población (t-educación, t-salud, t-gobierno, etc.) con énfasis en la inclusión social y digital.

2- Crear el “Foro Iberoamericano de TVDi” para incentivar el debate y la investigación conjunta. En particular,

crear un portal web para el foro que contenga:

- Información relativa a la aparición de nueva tecnología en el área temática de la TV digital Interactiva, incluyendo comentarios de profesionales e investigadores de la comunidad iberoamericana.

- Estudios sobre implantación y test específicos realizados tanto por los grupos de la red como por otros de la comunidad iberoamericana. Incluyendo llamadas a la participación.

- Difusión de las convocatorias y eventos de cooperación en el tema TVDi

- Búsqueda de socios para proyectos en cooperación que puedan ser de interés para la comunidad Iberoamericana en temas relacionados con la TVDi.

- Información sobre cursos y acciones formativas en el tema.

3- Realizar un evento anual de difusión abierto a la comunidad donde se ofrezcan charlas, por parte de los investigadores integrantes de la RedAUTI, relacionadas con el desarrollo de aplicaciones interactivas y contenidos para TVDi, estimulando la cultura local y el interés por innovación tecnológica entre los jóvenes;

4- Ofrecer cursos por parte de los investigadores integrantes de la RedAUTI en temas relacionados con el desarrollo de aplicaciones interactivas y contenidos para TVDi.

5- Incentivar las estancias de trabajo de doctorandos e investigadores en las instituciones participantes de países que se encuentren en estadio más adelantado de aplicaciones y desarrollo de contenidos.

7- Avanzar en la investigación transdisciplinaria para el desarrollo de aplicaciones, servicios y producción de contenidos sobre TVDi, creando grupos de trabajo y asociaciones estratégicas de investigación para buscar la creación de nuevos servicios en sectores emergentes

como son por ejemplo los juegos y la educación, la TV social, inteligencia ambiental, etc.

Entre los principales resultados de la RedAUTI se resaltan:

- Difusión mediante el portal web y redes sociales con participación de los miembros enviando información para difusión como publicaciones, proyectos, noticias, eventos, etc.

- Organización de un evento científico anual denominado Jornadas de difusión y capacitación de Aplicaciones y Usabilidad de la TVDi, habiéndose realizado jAUTI2012, jAUTI2013, jAUTI2014 y jAUTI2015 con la participación de la mayoría de grupos de la RedAUTI y participantes externos, con un total aproximado de 100 ponencias en los últimos cuatro años

- Publicación conjunta de 6 libros, cada uno con aproximadamente 20-25 artículos [Aba13][Aba14][Aba15a][Aba15b][Bib16] y una decena de publicaciones conjuntas en congresos internacionales/revistas

- Financiación de aproximadamente 80 viajes en cuatro años de proyecto para asistencia a los eventos científicos de la RedAUTI.

- Financiación de más de 10 estancias de investigación en instituciones de la red.

- Dictado de una veintena de cursos presenciales a lo largo de cuatro años y un curso masivo on-line sobre "Introducción a la TVDi".

- Avances en la propuesta de una maestría conjunta sobre TVDi a dictarse en la Universidad de Andalucía.

- Presentación de productos y servicios al sector empresas en los eventos relacionados por la RedAUTI y otros organismos como AR-SAT y la Comisión Nacional de Comunicaciones

Formación de recursos humanos

La formación de recursos humanos es prioritaria dentro de la RedAUTI. Los cursos presenciales dictados durante los cuatro años que lleva el proyecto tuvieron como destinatarios a un total de 500 alumnos. El curso masivo tuvo un total de 800 inscriptos, de los cuales aproximadamente 100 concluyeron el curso presentando la totalidad de actividades.

En las instituciones argentinas, miembros de la RedAUTI, hay en la actualidad algunas tesis en el marco de esta línea de investigación que comienzan a desarrollarse tal como:

- Producto de Software para aplicaciones TVDi basado en patrones de diseño. Alumna: Mirtha Miranda (Laboratorio de Tv Digital UNPA).

Existen otras que están más en sus inicios, pero que constituyen oportunidades de crecimiento para esta temática.

Referencias

[Aba13] Abásolo, M.J., De Giusti A. (Eds.) "Anales jAUTI 2012 I Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva" celebrado en La Plata (Argentina) en octubre 2013. ISBN: 978-950-34-0945-9. Universidad Nacional de La Plata (Argentina), 2013. Disponible on-line en <http://sedici.unlp.edu.ar/handle/10915/25926>

[Aba14] Abásolo, M.J., Castro, C. de (Eds.) "Anales jAUTI 2013 II Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva" celebrado en Córdoba (España) en septiembre 2013, ISBN: 978-84-697-0302-1. Universidad de Córdoba (España), 2014. Disponible on-line en <http://sedici.unlp.edu.ar/handle/10915/35693>

- [Aba15a] Abásolo, M.J., Kulesza, R. (Eds.) *“Proceedings de jAUTI 2014 III Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva | Workshop WTVI Webmedia 2014”* celebrado en Joao Pessoa, Paraíba (Brasil), Noviembre 2014, ISBN: 978-950-34-1188-9, Universidad Nacional de La Plata (Argentina), 2015. Disponible on-line en <http://sedici.unlp.edu.ar/handle/10915/44899>
- [Aba15b] Abásolo, M.J., Kulesza, R. (Eds.) *“Applications and Usability of Interactive TV. Third Iberoamerican Conference, jAUTI 2014, and Third Workshop on Interactive Digital TV, Held as Part of Webmedia 2014, João Pessoa, PB, Brazil, November 18-21, 2014. Revised Selected Papers”* ISBN: 978-3-319-22656-9, Communications in Computer and Information Science Springer-Verlag, 2015 <http://www.springer.com/us/book/9783319226552>
- [Art15] Artola, V., Sanz, C., Abásolo, M.J. *“Experiencias de TVDI en Argentina”* en A.Bibiloni, F. Perales, M.J.Abásolo (Eds.) *“Proceedings del VI Congreso de Televisión Digital Interactiva - jAUTI 2015 IV Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva”* celebrado en Palma (España) del 14 al 16 de octubre 2015, ISBN: 978-84-608-5567-5, Universitat de les Illes Balears (España), 2016
- [Bib16] A.Bibiloni, F. Perales, M.J.Abásolo (Eds.) *“Proceedings del VI Congreso de Televisión Digital Interactiva - jAUTI 2015 IV Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva”* celebrado en Palma (España) del 14 al 16 de octubre 2015, ISBN: 978-84-608-5567-5, Universitat de les Illes Balears (España), 2016
- [Cas14] Casas S., Oyarzo F. y Herrera F. *“API TVD, un Generador de Aplicaciones Interactivas para TV Digital”* Simposio Latinoamericano de Ingeniería de Software CLEI 2014, pp 298-305, ISBN 978-1-4799-6129-0, Montevideo, Uruguay, 2014
- [Cas15] Sandra Casas *“Herramientas para el Desarrollo de Aplicaciones TVDI a Escala y de Calidad”* 3er. Foro Internacional de Televisión Digital de La Habana (FDTV 2015), Cuba, 14-16 de Septiembre, 2015
- [Gon13] Nestor Daniel González *“Contenidos en la TV digital argentina. Estrategias y actores”* en Abásolo, M.J., De Giusti A. *“Anales jAUTI 2012 I Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva”* celebrado en La Plata (Argentina) en octubre 2013. ISBN: 978-950-34-0945-9. Universidad Nacional de La Plata (Argentina), 2013.
- [Gon14] Nestor Daniel González *“La interactividad en la Televisión Digital en Argentina. Desafíos y perspectivas”* en Abásolo, M.J., Castro, C. de *“Anales jAUTI 2013 II Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva”*celebrado en Córdoba (España) en septiembre 2013, ISBN: 978-84-697-0302-1. Universidad de Córdoba (España), 2014
- [Gon15] Nestor Daniel González *“Licencias de televisión digital. Tensiones entre viejos y nuevos actores”* en Abásolo, M.J., Kulesza, R. *“Proceedings de jAUTI 2014 III Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva | Workshop WTVI Webmedia 2014”* celebrado en Joao Pessoa, Paraíba (Brasil), Noviembre 2014, ISBN: 978-950-34-1188-9, Universidad Nacional de La Plata (Argentina), 2015.
- [Mar16] Sabrina Martorelli, Sergio Martorelli, Cecilia Sanz, María José Abásolo, Patricia Pesado *“HistoTV: Primeros pasos hacia una aplicación educativa para TVDi sobre Histología Animal”* en A.Bibiloni, F. Perales, M.J.Abásolo (Eds.) *“Proceedings del VI Congreso de Televisión Digital Interactiva - jAUTI 2015 IV Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva”* celebrado en Palma (España) del 14 al 16 de octubre 2015, ISBN: 978-84-608-5567-5, Universitat de les Illes Balears (España), 2016

[Mir16] Mirtha F.Miranda, Sandra Casas “*Desarrollo de una Línea de Producto de Software para aplicaciones de TVDi*” en A.Bibiloni, F. Perales, M.J.Abásolo (Eds.) “Proceedings del VI Congreso de Televisión Digital Interactiva - jAUTI 2015 IV Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva” celebrado en Palma (España) del 14 al 16 de octubre 2015, ISBN: 978-84-608-5567-5, Universitat de les Illes Balears (España), 2016

[San13] Graciela Santos, Andrea Miranda “*Interacciones en procesos educativos con tecnología. Algunas consideraciones para TVDi*” en Abásolo, M.J., De Giusti A. (Eds.) “Anales jAUTI 2012 I Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva” celebrado en La Plata (Argentina) en octubre 2013. ISBN: 978-950-34-0945-9. Universidad Nacional de La Plata (Argentina), 2013

[San16] Graciela Santos, Andrea Miranda y Silvia Stipcich “*Interactividad y Modelo Pedagógico. Configuración de una metodología de diseño y análisis*” en A.Bibiloni, F. Perales, M.J.Abásolo (Eds.) “Proceedings del VI Congreso de Televisión Digital Interactiva - jAUTI 2015 IV Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva” celebrado en Palma (España) del 14 al 16 de octubre 2015, ISBN: 978-84-608-5567-5, Universitat de les Illes Balears (España), 2016

[Zan13] Alejandra Zangara, Cecilia Sanz “*Aproximaciones al concepto de interactividad educativa*” en en Abásolo, M.J., De Giusti A. (Eds.) “Anales jAUTI 2012 I Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva” celebrado en La Plata (Argentina) en octubre 2013. ISBN: 978-950-34-0945-9. Universidad Nacional de La Plata (Argentina), 2013

[Zan14] Alejandra Zangara, Cecilia Sanz, María José Abásolo “*Modelos de EAD y producción de materiales educativos. El caso de la TVDi*” en Abásolo, M.J., Castro, C. de (Eds.) “Anales jAUTI 2013 II Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva” celebrado en Córdoba (España) en septiembre 2013, ISBN: 978-84-697-0302-1. Universidad de Córdoba (España), 2014.

Selección y manipulación de Recursos Educativos Abiertos en el entorno educativo argentino mediante aplicaciones basadas en ontologías

Sebastián Guaraz⁽¹⁾, Alejandro Sánchez⁽²⁾, Germán Montejano⁽²⁾⁽³⁾

⁽¹⁾ Facultad de Tecnología y Ciencias Aplicadas
Universidad Nacional de Catamarca
Maximio Victoria 55 - San Fernando del Valle de Catamarca
Tel. +54 -0383- 4435112
saguaraz10@gmail.com

⁽²⁾ Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis
Ejercito de Los Andes 950 - (5700) - San Luis - San Luis - Argentina
Tel: +54-266- 4520300 - Int 2128
{asanchez,gmonte}@unsl.edu.ar,

⁽³⁾ Facultad de Ingeniería, Facultad de Ciencias Exactas y Naturales
Universidad Nacional de la Pampa
Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina
Tel.: +54-2954-245220 - Int. 7125

Resumen

Los Recursos Educativos Abiertos (REAs) son materiales en formato digital que típicamente se ofrecen de manera gratuita y abierta para educadores, estudiantes y autodidactas para su uso, reutilización y modificación en la enseñanza, el aprendizaje y la investigación. A pesar de que existe un gran número de REAs disponibles en la web, a los docentes les resulta difícil su selección y adopción.

Esta línea de investigación, desarrollo e innovación (I-D-I) está centrada principalmente en facilitar al docente de nuestro país, la selección y manipulación de REAs a través de aplicaciones basadas en ontologías. De esta forma se podrá obtener recursos que se correspondan con ciertos parámetros, de acuerdo a una

determinada licencia de uso. Se podrá hacer uso del mismo, o se puede combinar con otro material, para luego ser compartido. Deberá permitir a los docentes organizar y describir REAs en términos de las ontologías subyacentes, sin ser ellos expertos en lenguajes de especificación de ontologías.

Palabras clave: Recursos Educativos Abiertos, licencias, Ontologías.

Contexto

Esta línea de I-D-I se enmarca en la convocatoria de Proyectos de Investigación y Desarrollo 2016 de la Facultad de Tecnología y Cs. Aplicadas de la Universidad Nacional de Catamarca y es una Línea Prioritaria para esta unidad académica según Resolución N° 258-15: El conocimiento científico y las

tecnologías de la información y las comunicaciones aplicadas a los procesos de enseñanza-aprendizaje para la mejora de la calidad educativa.

Además está vinculado con el proyecto “Ingeniería de Software”¹ de la Facultad de Ciencias Físico-Matemáticas y Naturales de la Universidad Nacional de San Luis (UNSL), y que incluye acciones de cooperación con otras universidades nacionales y del exterior en Europa y en América.

Introducción

Los docentes a menudo deben elaborar su plan de clases con sus contenidos conceptuales, de acuerdo a un proyecto áulico propuesto para el dictado de su materia. Para tal fin necesita de un conjunto de recursos pedagógicos y didácticos que amplíen y reafirmen los conocimientos sobre la temática deseada. En su búsqueda de materiales para la concreción de su proyecto áulico, que aporten valor al proceso de enseñanza-aprendizaje, el docente recurre a bibliografía impresa proporcionada por bibliotecas públicas, o a la gran diversidad de recursos disponibles en Internet.

Sin embargo, cada recurso educativo se encuentra enmarcado dentro de un conjunto de derechos llamado propiedad intelectual, que otorga privilegios a cada autor, sobre la creación y prestaciones que se realicen sobre sus obras.

El copyright es un tipo de licencia privativa que permite la comercialización de las creaciones, impidiendo su copia y distribución sin la obtención de los permisos correspondientes. Los recursos educativos bajo licencia copyright, tienen un costo asociado, se encuentran

distribuidos en internet dificultando su rápida obtención, no se permite realizar copias de esos recursos con fines de lucro, o para su difusión en público, y cualquier referencia de su contenido debe citar los autores. Estas características de las licencias privativas representan un condicionante para los docentes y alumnos, que deben afrontar el desarrollo de una clase que aporte un aprendizaje significativo con contenidos variados y de calidad.

Siguiendo la filosofía del software libre como tendencia mundial, surgieron otros tipos de licencias sobre documentos escritos o digitales que dan libertades de uso a sus destinatarios. Pueden ser adquiridas por lo general de forma gratuita, se pueden modificar o mezclar con otros recursos, o pueden ser compartidos bajo las mismas condiciones de usabilidad. Estas condiciones de libre acceso, no hacen que el autor pierda sus derechos, debiendo ser debidamente citadas las producciones utilizadas para otro fin. Entre ellas la más utilizada es la Licencia Creative Commons.

Los REAs están basados en este tipo de licencias, son materiales en formato digital que se ofrecen de manera gratuita y abierta.

Según la clasificación de (Valverde 2010), los REAs incluyen:

- Contenidos de aprendizaje: cursos completos, materiales para cursos, módulos, objetos de aprendizaje, colecciones y revistas (p.ej. OCW).
- Herramientas tecnológicas: software para la creación, entrega, uso y mejora del contenido de aprendizaje abierto, como los Sistemas de Gestión de Contenidos o CMS (p.ej. EduCommons), los Sistemas de Gestión de Aprendizaje o LMS (p.ej. Moodle), software para la producción colaborativa de conocimiento (p.ej. mediawiki.org) o herramientas para

¹ <http://www.sel.unsl.edu.ar/pro/proyec/2012/index.html>

desarrollar y publicar contenido abierto (p.ej. www.cnx.org).

- Recursos de implementación: licencias de propiedad intelectual que promuevan la publicación abierta de materiales (p.ej. Creative Commons), principios de diseño y adaptación local de contenido.

El origen del movimiento REA se remonta a los años 2000, para luego ganar visibilidad cuando el Instituto de Tecnología de Massachusetts (MIT) lanzó en 2001 su programa de recursos abiertos a gran escala, Open Course Ware. El término REA (Recurso Educativo Abierto, o OER, del inglés, Open Educational Resource) se utilizó por primera vez en el 2002, en el Foro de la UNESCO sobre el impacto del material educativo abierto en la educación superior.

Los REAs tienen el potencial de permitir a los ciudadanos de todo el mundo acceder a la educación desde cualquier lugar y en cualquier momento.

La comunidad educativa puede verse favorecida por este nuevo movimiento de recursos abiertos que se diversifica a nivel mundial. Pero esta diversificación trae nuevos inconvenientes que los docentes deben afrontar. Un gran número de autores, temáticas, contenidos no calificados, gran cantidad de repositorios, necesidad de conocimientos específicos sobre el uso de Tecnologías de la Información y la Comunicación (TIC) u organizacionales, entre otros factores, dificultan la selección.

Basado en la búsqueda y observación de algunos sitios web, portales o repositorios digitales se reconoce el trabajo y el esfuerzo realizado por distintas Instituciones Educativas, o por Organismos Gubernamentales que buscan la difusión de REA. A modo de ejemplo, describimos un grupo de ellos.

El portal educativo del Ministerio de Educación de la República Argentina, reúne gran cantidad de material clasificado por niveles de enseñanza, tipos de recursos y formatos.

La biblioteca electrónica de ciencia y tecnología del Ministerio de Ciencia, Tecnología e Innovación Productiva, incorpora enlaces de los portales que brindan acceso a revistas científico-técnicas, tesis, informes de investigación, presentaciones a congresos y documentación científica de producción nacional y de acceso abierto.

El Sistema Nacional de Repositorios promueve el acceso abierto y el intercambio de la producción científico-tecnológica generada en el país. Además, las instituciones adheridas pueden solicitar financiamiento para crear/fortalecer sus repositorios digitales y para la formación de recursos humanos.

La Biblioteca Digital de la Universidad Católica Argentina es un repositorio institucional que alberga tesis y trabajos finales seleccionados por cada Facultad, revistas y documentos de investigación, ponencias presentadas en jornadas, libros, etc.

El Sistema de Bibliotecas y de Información Secretaría de Ciencia y Técnica – Universidad de Buenos Aires.

En general, estos repositorios utilizan criterios sintácticos para retornar resultados, que se obtienen a partir de descripciones en lenguaje natural y/o alguna taxonomía ad-hoc. Se produce un número irrelevante de resultados que desalientan al docente en la selección y adopción de REA

De hecho, los docentes requieren contar con conocimientos en dominios ajenos a la docencia para una integración de REAs en su entorno educativo, como ser: los distintos tipos de licencias, derechos de autor y de propiedad, repositorios de REA (y sus taxonomías

cuando existen), formatos y mecanismos para acceder a esos recursos, involucrando la creación, modificación y difusión de contenidos.

Un enfoque posible es la selección de recursos a partir de su descripción expresada en términos de ontologías (Borst 1997, Gruber 1993). Existen fundamentos semánticos sólidos, como las lógicas descriptivas (Baader 2003), y lenguajes y tecnologías que soportan este enfoque (Hertel 2009). Un motor de inferencia clasifica los recursos que satisfacen una descripción conceptual de lo buscado especificada de acuerdo a la misma ontología con que se anotaron los recursos.

Existen diferentes tipos de ontologías que deben ser tenidos en cuenta (Guarino 1998). Las ontologías de dominio son uno de los ejes en este tipo de sistemas. En el caso de REAs, una ontología para las licencias de contenidos educativos debe considerarse. Estas ontologías seguramente utilicen conceptos definidos en ontologías de alto nivel, como ser, por ejemplo, la ontología de unidades estándar de medida. A su vez, aplicaciones brindando servicios más específicos utilizarán ontologías de aplicación que incorporan los conceptos definidos en ontologías de dominio. Las correspondencias en esta red de ontologías es especificada con tecnologías de “ontology matching” (Euzenat 2013).

El desarrollo de las ontologías se produce en el marco de metodologías existentes que se ajustan a distintos posibles contextos. En muchos casos, estas metodologías facilitan la colaboración de expertos de dominio, no familiarizados con lenguajes de especificación de ontologías (Grüninger y Fox 1995), (Uschold y King 1996), (Bernaras, et all 1996), (Gómez-Pérez et

all 1996), (Staab et al 2001), (Corcho et all 2003), (Tautz y Wangenheim 1998).

Líneas de Investigación, Desarrollo e Innovación

Esta línea de I-D-I se inicia persiguiendo la noción de facilitar al docente la selección y manipulación de REA. Se plantea la pregunta si es posible emplear fundamentos, metodologías y tecnologías de explicitación del conocimiento para acercar las diferencias conceptuales y de lenguaje entre quienes desarrollan, publican y (re)utilizan REA en el contexto educativo argentino (extensivo al hispano-parlante). Los principales ejes de la línea son:

- Participación de expertos del dominio de la región como futuros usuarios y fuente de conocimiento.
- Selección y customización de metodología.
- Utilización de herramientas y lenguajes basados en lógicas descriptivas.

Objetivos y resultados esperados

Se plantea como objetivo general facilitar al docente la selección y manipulación de REAs a partir de aplicaciones basadas en ontologías. De este objetivo general se desprenden actividades que se espera generen el conjunto de resultados que sigue:

- Descripción informal de la conceptualización de los problemas y dominios relacionados.
- Una red de ontologías, (re)utilizando ontologías existentes, que provean una vista global sobre repositorios de REA.
- Aplicaciones basadas en ontologías que permitan al docente seleccionar REAs, ayudando a superar las

dificultades conceptuales y lingüísticas que esto implica.

- Aplicaciones basadas en ontologías que permitan anotar REA a quienes los desarrollan, publican y (re)utilizan.

Formación de Recursos Humanos

La siguiente línea de I-D-I es el punto de partida para la formación de recursos humanos mediante el desarrollo de tesis de postgrado y tesinas. Se prevé continuar formando a integrantes del proyecto de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca. Integrantes del proyecto desarrollarán su Tesis de Maestría en Ingeniería de Software y alumnos de grado completarán sus tesinas dentro de la línea de I-D-I. Es importante notar el apoyo académico y científico para los temas desarrollados en la formación de recursos humanos calificados ya que son carreras de posgrado categorizadas "A" para los desarrollos de tesis de posgrado, y carreras de grado acreditadas por 6 años para los desarrollos de las tesinas.

Referencias

Alice Hertel, Jeen Broekstra, Heiner Stuckenschmidt (2009). RDF Storage and Retrieval Systems. Handbook on Ontologies. Part of the series International Handbooks on Information Systems pp 489-508.

Amorim, R. R., Rabelo, T., & Amorim, D. (2012). Open Educational Resources Ontology. In Anais dos Workshops do Congresso Brasileiro de Informática na Educação (Vol. 1, No. 1).

Bernaras, A., Laresgoiti, I., & Corera, J. (1996). Building and Reusing Ontologies for Electrical Network

Applications'. In ECAI (pp. 298-302). PITMAN.

Berrocoso, J. V. (2010). El movimiento de "educación abierta" y la "universidad expandida". *Tendencias pedagógicas*, (16), 157-180.

Borst, Willem Nico (1997). "Construction of Engineering Ontologies for Knowledge Sharing and Reuse". Tesis doct. Enschede.

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point?. *Data & knowledge engineering*, 46(1), 41-64.

Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider (Eds.). 2003. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, New York, NY, USA.

Gómez-Pérez, A., Fernández, M., & Vicente, A. D. (1996). Towards a method to conceptualize domain ontologies.

Guarino, N. (1998). Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy. 1st. Amsterdam, The Netherlands, The Netherlands: IOS Press.

Gruber, Thomas R. (1993). "A Translation Approach to Portable Ontology Specifications". En: *Knowl. Acquis.* 5.2, págs. 199-220.

Grüninger, M., & Fox, M. S. (1995). *Methodology for the Design and Evaluation of Ontologies*.

Piedra, N., Chicaiza, J., López, J., Tovar, E., & Martínez, O. (2010, July). Design study of OER-CC Ontology-A semantic web approach to describe open educational resources. In *Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on* (pp. 207-209). IEEE.

Paradigmas de Interacción Persona Ordenador en el ámbito de la Educación y la Educación Especial. Avances del proyecto y resultados

Sanz Cecilia¹, Moralejo Lucrecia¹, Artola Verónica^{1,5}, Salazar Mesía Natalí¹, Guisen Andrea⁶, Baldassarri Sandra², Manresa Yee Cristina⁴, Pesado Patricia^{1,3}

¹ Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata

² GIGA AffectiveLab, Universidad of Zaragoza, España

³ Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)
526 e/ 10 y 11 La Plata Buenos Aires

⁴ Departamento de Ciencias Matemáticas e Informática de la Computación. Universidad de Islas Baleares, España

⁵ Becaria doctoral CONICET

⁶ Becaria post-doctoral CONICET

{[csanz](mailto:csanz@unlp.edu.ar), [lmoralejo](mailto:lmoralejo@unlp.edu.ar), [variola](mailto:variola@unlp.edu.ar), [nsalazar](mailto:nsalazar@unlp.edu.ar), [aguisen](mailto:aguisen@unlp.edu.ar), [ppesado](mailto:ppesado@unlp.edu.ar)}@lidi.info.unlp.edu.ar, sandra@unizar.es,
cristina.manresa@uib.es

CONTEXTO

Esta investigación forma parte del proyecto “Tecnologías para Sistemas de Software Distribuidos. Calidad En Sistemas y Procesos. Escenarios Educativos Mediados Por TIC (período 2014-2017)”, perteneciente al Instituto de Investigación en Informática LIDI, de la Facultad de Informática de la UNLP y acreditado por el Ministerio de Educación de la Nación.

RESUMEN

El subproyecto “Escenarios educativos mediados por Tecnologías de la Información y la Comunicación (TIC)” forma parte del proyecto mencionado en la sección previa.

Se presenta en este trabajo uno de los ejes temáticos que forma parte de dicho sub-proyecto, referido a Paradigmas de Interacción Persona Ordenador y sus posibilidades específicas para el ámbito educativo, y en particular, para educación especial. Se describen las líneas de investigación en las que se trabaja y los resultados alcanzados durante el 2015 e inicios de 2016.

Palabras claves: Tecnologías de la Información y la Comunicación, Realidad Aumentada, Interacción Tangible, Computación Afectiva.

INTRODUCCION

La interacción entre personas y ordenadores ha evolucionado en los últimos años y ha tomado auge su investigación. Hoy en día se busca crear nuevas interfaces adaptables a los dispositivos, al entorno y a los usuarios (Bertini, 2004).

Uno de los temas de interés en esta línea de investigación es la Interacción Tangible (IT) en educación. La manipulación física de objetos ha sido la base del desarrollo cognitivo y psicomotriz en la educación preescolar y también, resulta efectiva en adultos, y en educación especial (Marcos, Baldassarri & Cerezo, 2012; Sanz, Baldassarri, Guisen, Marco, Cerezo & De Giusti, 2012; Piper, O'Brien, Morris & Winogrand, 2006). Según trabajos como el de (O'Malley & Fraser, 2004), el de (Marshall, 2007) y el de (Dillenbourg & Evans, 2009), las posibilidades de IT en el ámbito educativo son sustanciales, especialmente para aquellas actividades que involucran la colaboración entre personas. Según McNeil y Jarvin (2007) trabajar con interfaces tangibles proporciona un canal adicional para transmitir información, activan el conocimiento del mundo real y mejoran la memoria a través de la manipulación y la acción física que mantienen con el objeto.

Estas cualidades se ponen en evidencia en las experiencias como las presentadas en (Fernaes y Tholander, 2005; Rick, Harris, Marshall, Fleck, Yuil & Rogers, 2009; Karrufa, Leat & Olivier, 2010; Spermon, Schoute & Hoven, 2014).

Por otra parte, la realidad aumentada (RA) es otro paradigma de interacción persona ordenador que ha ido creciendo en interés.

Las aplicaciones con RA pueden potenciar habilidades tales como la capacidad espacial, las habilidades prácticas y la comprensión conceptual. Además, los entornos de RA pueden motivar a los estudiantes y crear posibilidades de aprendizaje colaborativo en entornos no tradicionales con contenidos virtuales (Cubillo Arribas, Martín Gutiérrez, Castro Gil & Colmenar Santos, 2014). Al mismo tiempo, la interacción tangible y la realidad aumentada empleadas en escenarios de educación especial presentan interesantes posibilidades, que son expuestas en trabajos como los de (Sanz, Guisen, De Giusti, Baldassarri, Marco & Cerezo, 2013; Gonzalez, Martínez, Villanueva, Vallejo & López, 2011; Garzotto & Bordogna, 201).

Por otra parte, la computación afectiva es otra línea de investigación actual. Los seres humanos son eminentemente emocionales, y la base de su interacción social se vincula con la capacidad de comunicar sus emociones y percibir la de los otros. El principal objetivo de la computación afectiva es capturar y procesar la información afectiva de manera tal de mejorar la comunicación entre las personas y los ordenadores (López, Cearreta, Garay-Vitoria, López de Ipiña & Beristain, 2009). En los entornos virtuales de enseñanza y aprendizaje donde los alumnos y docentes participan y se comunican como parte del proceso educativo, se vuelve fundamental mejorar la comunicación mediada a través del ordenador, de manera tal de poder percibir las emociones del otro. Es por ello que resulta de interés abordar esta línea de investigación en el marco de este proyecto. En los trabajos de Qi Luo (2006) y Lan Li, Li Cheng & Kun-xi Qian (2008) se analizan los beneficios de los entornos virtuales de enseñanza y aprendizaje con posibilidades de computación afectiva. En otros trabajos como el de (Garay, Cerraeta, López y Fajardo, 2006) se aborda la importancia de la computación afectiva en el ámbito de la comunicación en personas con necesidades especiales. En el proyecto que aquí se presenta se ha iniciado la investigación y desarrollo de aplicaciones en estos temas y se han enfocado algunas tesis que se detallan en la sección de resultados.

LINEAS DE INVESTIGACION / DESARROLLO

Las líneas de investigación, desarrollo e innovación en las que se trabaja son

- Estilos y Paradigmas de Interacción Persona – Ordenador. Posibilidades para la educación, y con foco en la accesibilidad e inclusión.
- Realidad Virtual y Aumentada. Sus posibilidades para el ámbito educativo
- Interacción tangible y multitáctil. Aplicaciones en educación y educación especial
- Computación afectiva para el enriquecimiento de escenarios educativos
- Herramientas colaborativas para la educación y educación especial en particular.

RESULTADOS ESPERADOS/OBTENIDOS

En esta sección se detallan los resultados alcanzados durante el 2015 del proyecto aquí presentado, y otros que se espera alcanzar durante este año.

Así, en 2015, se ha avanzado en el desarrollo de una tesis doctoral orientada a la investigación de mundos virtuales para la enseñanza de personas con hipoacusia. En este sentido se está abordando el desarrollo de un espacio en OpenSim (motor para generar mundos virtuales) para llevar adelante un estudio de caso que se abordará durante el 2016 (Fachal, Sanz & Abásolo, 2015).

Se participó en el desarrollo de cuatro capítulos de un libro (en edición), en conjunto con la Universidad de Guadalajara en el marco del Programa Pablo Neruda de la Red *IberoTICs*. En estos capítulos se presentaron algunas experiencias vinculadas a diferentes paradigmas de interacción persona ordenador y sus posibilidades para el escenario educativo.

También se trabajó en la difusión de los avances realizados a partir de una tesis doctoral sobre el desarrollo de un entorno colaborativo para la comunicación aumentativa (Guisen & Sanz, 2014).

Se continuó con el desarrollo de aplicaciones de interacción tangible utilizando VisionAR, una mesa interactiva (*tabletop*), creada en el marco del proyecto y con la colaboración del grupo GIGA *Affective Lab* de la Universidad de Zaragoza.

En 2014 se diseñó e implementó EPIT una aplicación de interacción tangible para la Enseñanza de Programación (Artola, Sanz, Gorga & Pesado, 2014). Durante 2015 se han realizado sesiones con docentes que evaluaron la aplicación y, acorde a esto, se implementaron varias mejoras. Al mismo tiempo, se ha realizado una primera sesión de prueba con alumnos de la asignatura Programación I de la carrera Ingeniería en Computación. Se espera para este año continuar con las sesiones de prueba con el fin de validar si se alcanzan los objetivos educativos para los que fue creada EPIT.

También, se continuó con la aplicación de ITCol (un juego basado en interacción tangible para la colaboración). ITCol fue utilizado en el marco de un seminario de la Maestría en Tecnología Informática Aplicada en Educación de la Facultad de Informática de la UNLP, para que los alumnos vivencien una experiencia de trabajo colaborativo cara a cara mediado por tecnología digital. Se realizó una evaluación de estas sesiones a través de entrevistas y encuestas, y se utilizó registro filmico. Los resultados están siendo analizados y se presentarán en futuras publicaciones (Artola & Sanz, 2015).

Por otra parte, se está trabajando en una publicación en relación a las últimas sesiones de evaluación de ACoTI (*Augmentative Communication through Tangible Interaction*), esta aplicación consiste en una serie de juegos de asociación que permiten el trabajo con objetos reales y virtuales. Los juegos han sido planificados con el fin de apoyar el proceso de desarrollo del lenguaje y la comunicación en alumnos con necesidades complejas de comunicación (Speech Pathology Association of Australia Limited, 2014).

Se inició el desarrollo de una herramienta de autor para la construcción de actividades educativas basadas en interacción tangible. Se han estudiado diferentes editores de aplicaciones de interacción tangible que han servido de antecedente para el desarrollo de la herramienta de autor antes mencionada (Artola, Sanz, Moralejo, Baldassarri & Pesado, 2015).

En relación a la línea de realidad aumentada y virtual, se implementaron varias plantillas de la herramienta de autor AuthorAR, que forma parte de una tesis de maestría en Tecnología Informática Aplicada en Educación (TIAE) (Moralejo, Sanz & Pesado, 2014). Se realizaron también avances en el desarrollo de actividades educativas con realidad aumentada para la enseñanza de Programación, como parte de otra tesis de maestría en TIAE (Salazar Mesía, Gorga & Sanz, 2015).

Se ha presentado también una propuesta de tesis de doctorado: “Reconocimiento de información afectiva en Entornos Virtuales de Enseñanza y Aprendizaje, a partir de espacios de interacción textual”, que se vincula con la aplicación de técnicas de computación afectiva a los espacios de interacción textuales de los EVEA (Rodríguez, 2014). Esta es dirigida por miembros del proyecto, y se espera avanzar en el relevamiento de antecedentes durante el 2016.

Se está iniciando una tesis de maestría en TIAE relacionada al análisis de las interacciones de alumnos en los foros virtuales de cursos en entornos virtuales de enseñanza y aprendizaje para la construcción de mapas conceptuales que permitan sintetizar las principales ideas trabajadas.

Es importante destacar que el equipo de trabajo para esta investigación se integra por investigadores del III LIDI y otros de España (UIB y Zaragoza) que colaboran en el desarrollo de los temas y en la dirección de trabajos de tesis de maestría y doctorado.

En cuanto a los proyectos vinculados con la temática, el III- LIDI participa en los siguientes:

- Acuerdo de colaboración en estos temas con la Universidad de Zaragoza. En particular con el grupo GIGA *AffectiveLab*.
- Acuerdo de colaboración en estos temas con la Universidad de Islas Baleares.
- REDAUTI: Red Temática en Aplicaciones y Usabilidad de la Televisión Digital Interactiva. Compuesta por 170 investigadores de 28 grupos de investigación (22 universidades y 6 empresas) de 11 países iberoamericanos. Entidad financiadora: Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED). En el marco de este proyecto, se aborda la construcción de contenidos

para la TVDI que puedan ser de aporte también para el área de educación y educación especial.

- Acuerdo de cooperación con la Universidad Nacional de Santiago del Estero. Se asesora en temas relacionados con el m-learning.

FORMACION DE RECURSOS HUMANOS

Hay un becario de UNLP con beca TIPO A, que está realizando la tesis de maestría en la temática.

Un becario CONICET, que está realizando su doctorado en temas vinculados con Interacción Tangible para el escenario educativo. Un especialista en TIAE, que está finalizando su tesis de maestría.

Se está realizando otra tesis de doctorado vinculada a estos temas.

Se está iniciando un trabajo de grado también en esta línea de investigación.

En la siguiente sección se presenta: la bibliografía y los trabajos citados aquí, algunos textos de estudio que se utilizan en la investigación, y por otra parte, algunas publicaciones que forman parte de los resultados de este proyecto.

BIBLIOGRAFIA

- Artola V. & Sanz C. (2015). Informe técnico de las sesiones de evaluación realizadas con ITCOL.
- Artola, A., Sanz, C. V., Gorga, G. & Pesado P. (2014). Diseño de un juego basado en Interacción Tangible para la enseñanza de Programación. Argentina. San Justo, Buenos Aires, Argentina. Libro. Artículo Completo. Congreso. XX Congreso Argentino de Ciencias de la Computación (CACIC 2014).
- Artola, V., Sanz, C. V., Moralejo, L., Pesado, P. M., & Baldassarri, S. (2015). Herramienta de autor para la creación de actividades educativas basadas en Interacción Tangible. En XXI Congreso Argentino de Ciencias de la Computación. Junín, 2015.
- Battocchi A., Ben-Sasson A., Esposito G., Gal E. Pianesi F., Tomasini D., Venuti P., Weiss P. L. and Zancanaro M. (2010). "Collaborative Puzzle Game: a Tabletop Interface for Fostering Collaborative Skills in Children with Autism Spectrum Disorders". *Journal of Assistive Technologies*. 4(1): pp. 4-14
- Bertini, E. & Santucci, G. (2004). Modelling internet based applications for designing multi-device adaptative interfaces. Working conference on advanced visual interfaces, pp. 252-256.
- Boix J., Basil C. (2005). CAA en atención temprana. *Comunicación y pedagogía: NT y recursos didácticos*. ISSN: 1136-7733, N° 205, pp. 29-35.
- Bujak, K. R.; Radu, I.; Catrambone, R.; MacIntyre, B.; Zheng, R.; Golubski, G. (2013). A psychological perspective on augmented reality in the mathematics classroom. *Computers & Education*.
- Chen, C.; Tsai, Y. (2012). Interactive augmented reality system for enhancing library instruction in elementary schools. *Computers & Education*, 59 (2), (638652).
- Cubillo Arribas, J., Martín Gutiérrez, S., Castro Gil, M., Colmenar Santos, A. (2014). Recursos digitales autónomos mediante realidad aumentada. *RIED. Revista Iberoamericana de Educación a Distancia*, volumen 17, n° 2, pp. 241-274.
- Cuendet, S.; Bonnard, Q.; Do-Lenh, S.; Dillenbourg, P. (2013). Designing augmented reality for the classroom. *Computers & Education*.
- Dillenbourg, P. & Evans, M. (2011). Interactive tabletops in education. *Computer-Supported Collaborative Learning*. 6, pp. 491-514.
- Fachal A., Sanz C., Abásolo M.J. (2015). Informe técnico sobre "Antecedentes en el desarrollo de software y herramientas educativas para la enseñanza de personas con hipoacusia".
- Fernaeus, Y. & Tholander, J. (2005). Looking at the computer but doing it on land: children's interactions in a tangible programming space. In *Proc. of HCI*, 3-18.
- Garay, N., Cearreta, I., López, J. M. & Fajardo, I. Agora Center, University of Jyväskylä. DOI: <http://dx.doi.org/10.17011/ht/urn.2006159>
- Garzotto, F. & Bordogna, M. (2010). Paper-based multimedia interaction as learning tool for disabled children. *IDC '10: Proceedings of the 9th International Conference on Interaction Design and Children*.

- González, C., Martínez, M. A., Villanueva, F.J., Vallejo, D. & López, J. C. (2011). Sistema para la navegación en interiores mediante técnicas de Realidad Aumentada. Disponible en: <https://arco.esi.uclm.es/public/papers/2011-Ei3-carlos.gonzalez.pdf>. Recuperado en 2015
- Guisen A., Sanz C. (2014). Diseño De Ecca (Entorno Colaborativo De Comunicación Aumentativa Y Alternativa). Una Ayuda Tecnológica Para Alumnos Con Necesidades Complejas De Comunicación. Edutec-E. Revista Electrónica de Tecnología Educativa. ISSN: 1135-9250. Número 50. Fecha: Diciembre de 2014. URL: http://edutec.rediris.es/Revelec2/Revelec50/n50_Guisen_Sanz.html
- Guisen, A. & Sanz, C. (2014). Desarrollo de tecnologías educativas orientadas a la inclusión social de usuarios de sistemas de Comunicación Aumentativa asistidos. Revista IRICE. Vol.26. N.26., pp. 141 – 158. ISSN: 0327-392X. Disponible en: <http://web2.rosario-conicet.gov.ar/ojs/index.php/revistairice>
- Hsiao, K.; Rashvand, H. F. (2011). Body Language and Augmented Reality Learning Environment. 2011 Fifth FTRA International Conference on Multimedia and Ubiquitous Engineering. (246-250).
- Kharrufa, A., Leat, D. & Olivier, P. (2010). Digital Mysteries: Designing for Learning at the Tabletop? ITS10 Teaching & Learning. Saarbrücken, Alemania.
- Lan Li, Li Cheng & Kun-xi Qian. (2008). An e-Learning System Model Based on Affective Computing. International Conference on Cyberworlds 2008.
- Lin, Ch. & Chao, J.T. (2010). Augmented Reality-Based Assistive Technology for Handicapped Children. International Symposium on Computer, Communication, Control and Automation.
- López, J.M., Cearreta, I., Garay-Vitoria, N., López de Ipiña, K. & Beristain, A. (2009). A Methodological Approach for Building Multimodal Acted Affective Databases. In Engineering the User Interface. From Research to Practice. Ed. Springer. Editores: Redondo M., Bravo C. & Ortega M. ISBN:978-1-84800-135-0.
- Bojanovic, F., Manresa-Yee, C., Abásolo, M.J., Larrea, M. & Sanz, C (2015). Get Closer to Activate It! An Educational Tool for People with Multiple Disabilities. XXI Congreso Argentino de Ciencias de la Computación. Junín, 2015. ISBN: 978-987-3806-05-6. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/50238>
- Marco J., Baldassarri S., Cerezo S. (2012). Tangible interaction and tabletops: new horizons for children's games. International Journal of Arts and Technology. Vol. 5. Nro2.
- Marshall, P. (2007). Do tangible interfaces enhance learning? In Proc 1st International Conference on Tangible and embedded interaction, pp. 163-170. Baton Rouge, Louisiana.
- Matt Dunleavy, C. D.; Mitchell, R. (2012). Affordances and Limitations of Immersive Participatory. Augmented Reality Simulations for Teaching and Learning, 18 (1), (7-22).
- McNeil N. & Jarvin L. (2007). When Theories Don't Add Up: Disentangling the Manipulatives Debate. Theory Into Practice. Vol. 46, no. 4, pp. 309 - 316
- Milgram, P.; Takemura, H.; Utsumi, A.; Kishino, F. (1994). Augmented Reality: A class of displays on the reality-virtuality continuum. Telemanipulator and Telepresence Technologies, 2351, 11.
- Moralejo L., Sanz C., Pesado P., Baldassarri S. (2013). AuthorAR: Authoring Tool For Building Educational Activities Based On Augmented Reality. 2013 International Conference on Collaboration Technologies and Systems. San Diego, Estados Unidos. Páginas 503-507 - ISBN 978-1-4673-6404-1.
- Moralejo Lucrecia; Sanz Cecilia V.; Pesado Patricia; Baldassarri S. (2014). Análisis comparativo de Herramientas de Autor para la creación de actividades de Realidad Aumentada. Argentina. Chilecito, La Rioja, Argentina. 2014. Libro. Artículo Completo. Congreso. IX Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2014).
- Moralejo, L., Sanz, C. & Pesado, P. (2014). Análisis comparativo de herramientas de autor para la creación de actividades de Realidad Aumentada. Estudio de sus características

- específicas para el escenario educativo. Trabajo de Especialización en TIAE. Aprobado en Diciembre de 2014. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/43605>
- Muntaner J. (2010). De la Integración a la Inclusión: un nuevo modelo educativo. V Congreso Internacional de Tecnología Educativa y Atención a la Diversidad.
 - O'Malley, C. & Fraser D.S. (2004), Literature Review in Learning with Tangible Technologies. NESTA Futurelab.
 - Piper, A. M., O'Brien, E., Morris, M. R. & Winograd, T. (2006). SIDES: a cooperative tabletop computer game for social skills development. 20th Conference on Computer Supported Cooperative Work.
 - Qi Luo. Application of Affective Computing in e-Learning System. (2006). Open Education Research. Pp. 80-83.
 - Rick, J., Harris, A., Marshall, P., Fleck, R., Yuill, N., & Rogers, Y. (2009). Children designing together on a multitouch tabletop: An analysis of spatial orientation and user interactions. In Proceedings of IDC '09 (pp.106-114). New York: ACM Press.
 - Rodriguez, V., Sanz, C. & Baldassarri, S. (2014). Reconocimiento de información afectiva en EVEAs a partir de espacios de interacción textual. Propuesta de tesis de doctorado. Aprobada en Octubre de 2014. Expediente: 3300-007114/13-001
 - Salazar Mesía, N., Gorga, G. & Sanz, C. (2015). EPRA: herramienta para la enseñanza de conceptos básicos de programación utilizando realidad aumentada. X Congreso sobre Tecnología en Educación & Educación en Tecnología. Corrientes, 2015. ISBN: 978-950-656-154-3, 426-435. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/49071>
 - Sanz C., Baldassarri S., Guisen A., Marco J., Cerezo E. De Giusti A. (2012). "ACoTI: herramienta de interacción tangible para el desarrollo de competencias comunicacionales en usuarios de comunicación alternativa. Primeros resultados de su evaluación". VII Congreso de Tecnología en Educación y Educación en Tecnología (TEyET 2012). Junín, Buenos Aires, Argentina. Actas del Congreso - ISBN 978-987-28186-0-9. Págs. 226-233.
 - Sanz, C., Guisen, A., De Giusti, A., Baldassarri, S., Marco, J. & Cerezo, E. (2013) Games As Educational Strategy: A Case Of Tangible Interaction For Users Of Alternative And Augmentative Communication. 2013 International Conference on Collaboration Technologies and Systems. San Diego, Estados Unidos. Págs. 377-381. ISBN 978-1-4673-6404-1.
 - Speech Pathology Association of Australia Limited: "Augmentative and Alternative Communication". (2014) http://www.speechpathologyaustralia.org.au/library/Clinical_Guidelines/AAC.pdf
 - Spermon, M., Schoute, I. & Hoven, E. (2014). Designing interaction in digital tabletop games to support collaborative learning in children. International Journal Learning Technologies, 9(1), pp 3-24.
 - Wojciechowski, R.; Cellary, W. (2013). Evaluation of learners' attitude toward learning in ARIES augmented reality environments. Computers & Education

La tecnología como vehículo de articulación Nivel Medio / Universidad en la provincia de La Rioja

Fernanda Beatriz Carmona, Alberto Eduardo Riba, Fernando Emmanuel Frati, Claudia Isaia, Alejandro Cruz, Jorge Tejada Germán Torres, Matías Pérez

Departamento de Básicas y Aplicadas, Universidad Nacional de Chilecito
9 de Julio 22, Chilecito, La Rioja, Argentina

{fbcarmona, ariba, fefrati, cisaia, acruz, jtejada}@undec.edu.ar, germanfntorres@gmail.com, mataguper20@gmail.com

Resumen

Esta línea de I+D pretende fortalecer la articulación Universidad – Nivel Medio, estimulando el diseño de mecanismos que permitan alcanzar un diagnóstico compartido y planes de trabajo, destinados a disminuir la brecha que existe entre estos dos niveles.

Se propone una estrategia pedagógica que, utilizando la Robótica Educativa y la “enseñanza directa” como aprendizaje por descubrimiento guiado, pretende apoyar a los procesos de enseñanza y aprendizaje, a través del desarrollo de una plataforma de hardware de bajo costo que pueda ser utilizada en las Escuelas para despertar vocaciones tempranas vinculadas a las carreras tecnológicas.

Los temas abordados son transversales a varias áreas, como educación, algoritmos, lógica, programación y arquitectura de computadoras.

El desarrollo de esta línea conlleva, sin dudas, a la generación de otros proyectos relacionados con el uso de la tecnologías en la articulación Universidad - Nivel Medio considerando la amplia cobertura de la temática a desarrollar y la imperiosa necesidad de articulación existente entre estos niveles educativos, como así también, en la gestación de líneas de investigación relacionadas con nuevas estrategias didácticas en Robótica y

Educación, la utilización y programación de microcontroladores en otros ambientes de trabajo y la utilización de lenguajes interpretados.

Palabras clave: Educación, capacitación, TIC, enseñanza-aprendizaje, robótica educativa, aprendizaje experimental, aprendizaje inductivo, programación, juego.

Contexto

Esta línea surge conforme a una investigación anterior sobre “Estudio de la Brecha Digital respecto del uso y aprovechamiento de las Tecnologías de la Información y la Comunicación (TIC) en las escuelas de la Región Chilecito - Famatina” realizado en colaboración con el Instituto de Formación Docente de la localidad de Famatina; el proyecto “La robótica como introducción a la formación tecnológica” aprobado en el año 2013 por la Secretaría de Políticas Universitarias y la colaboración de otras instituciones del país y del extranjero a través de los proyectos “Red para la Integración de Universidades en el uso de TIC para la Inclusión en la Educación Superior” aprobado en la Séptima Convocatoria a Redes Internacionales, año 2013, y del proyecto “Mejora de la Enseñanza de las Ciencias” convocatoria

La Universidad y la Escuela Secundaria, tercera etapa año 2015, ambas convocatorias de la Subsecretaría de Gestión y Coordinación de Políticas Universitarias – Ministerio de Educación de la Nación.

Introducción

Con la constante evolución de las TIC en nuestros días se hace necesaria la incorporación, integración y aprovechamiento pedagógico de las Tecnologías de la Información y la Comunicación en el Sistema Educativo, ya que la sociedad actual se caracteriza por procesos de cambio permanente y por una demanda cada vez mayor de conocimientos sobre las nuevas tecnologías.

Esta realidad trajo aparejada la preocupación sobre el uso real de la tecnología. Sea cual sea el nivel de integración de las TIC en los centros educativos, los docentes y estudiantes necesitan una “alfabetización digital” y una actualización didáctica.

Es necesario fortalecer la articulación Universidad / Nivel Medio, planteando nuevos escenarios donde los estudiantes no solo sean usuarios de las tecnologías sino participantes activos, generando una adecuada planificación y realización conjunta de acciones que favorezcan la inserción de los estudiantes en la Universidad para prevenir y disminuir las causas de deserción y estancamiento en los primeros años de las carreras universitarias.

La robótica como introducción a la formación tecnológica

Los procesos de desarticulación y segmentación educativa, desarrollados en los últimos años en nuestro país y en especial en la región, dificultan el pasaje, ingreso y permanencia, de los alumnos del nivel medio a los estudios

universitarios[1]. Las instituciones universitarias requieren de sus alumnos no solo los aprendizajes vinculados al conocimiento de las disciplinas, sino también los relacionados con las destrezas y estilos de pensamiento de las culturas disciplinares, y el desarrollo de una personalidad autónoma y crítica para su desempeño social [2].

Con el fin de acortar la brecha existente en la articulación Universidad / Nivel Medio se deben profundizar contenidos teóricos - prácticos y su aplicación en un proceso de formación integral que reúna conocimientos, habilidades, destrezas y actitudes propios de los modos de producción en los diferentes campos disciplinares, acordes a los requerimientos sociales.

En este contexto, existen políticas de estado destinadas a facilitar a todos los docentes y alumnos del Nivel Medio el acceso a la tecnología de la información y las comunicaciones: la implementación del programa del Estado Nacional denominado Modelo 1 a 1 a través del programa “Conectar Igualdad” al cual la provincia de La Rioja complementó con el plan “Joaquín V. González”, han alcanzado una alta cobertura, cumpliendo con la entrega de netbooks a estudiantes y docentes de escuelas secundarias. A estos programas se suma la iniciativa del Estado Nacional “program.ar 2020”, lanzada en el mes de Abril de 2014, que busca acercar a los jóvenes en edad escolar al aprendizaje de las Ciencias de la Computación y concientizar a la sociedad en general sobre la importancia de conocer estos conceptos.

Conforme a estas tendencias, consideramos oportuna la intervención de la UNDeC como nexo de articulación entre Nivel Medio y Universidad, dado que la universidad es el medio adecuado y pertinente para fortalecer los procesos de enseñanza y aprendizaje sobre dicha

tecnología, con el objetivo final de despertar en los estudiantes de Nivel Medio la vocación por las carreras tecnológicas, particularmente en Informática.

La Robótica Educativa se concibe como un contexto de aprendizaje que se apoya en las tecnologías digitales y en los procesos de mediación pedagógica para que los estudiantes creen prototipos o simulaciones robóticas que surgen a partir del ingenio, la creatividad y puesta en práctica de lo aprendido [3]. Es de interés plantear a la robótica como vehículo de aprendizaje con materiales concretos, motivando a los jóvenes a construir, diseñar y explorar nuevas formas de hacer las cosas, a través del aprendizaje experimental, el trabajo en equipo y el desarrollo de su confianza y habilidades innovadoras, brindando un espacio que les permita no sólo ser usuarios de las tecnologías, sino que, a partir de conocimientos matemáticos, mecánicos, físicos y lógicos, logren resolver, en forma activa, problemas significativos.

En Argentina y en Latinoamérica se están implementando proyectos, como propuestas de enseñanza de la programación en los primeros años de la carrera, que incorporan entre sus estrategias el aprendizaje basado en problemas y el aprendizaje colaborativo, otros incluyen, además, a la robótica como una opción para la profundización y gestación de habilidades cognitivo-creativas [5, 6, 7].

Algunas de estas propuestas promueven la construcción de robots que compiten de acuerdo a reglas internacionales y por categorías, otras usan la robótica como recurso de apoyo en el estudio de habilidades básicas en matemáticas, ciencias o física y construyen y programan modelos que ayudan a representar con elementos externos esos conceptos. Otros promueven la

construcción de robots que ejecutan tareas y funciones particulares o que se comportan de cierta manera ante variables del ambiente [8, 9].

Aunque todas ellas sirven de inspiración, aplicamos un enfoque diferente: utilizar actividades colaborativas y lúdicas relacionadas con la programación de un robot para la resolución de problemas, como estrategia de enseñanza de la programación destinada especialmente a estudiantes del Nivel Medio [10].

Se privilegia el aprendizaje inductivo por conocimiento guiado proponiendo:

- Generar un espacio de articulación Universidad - Nivel Medio que permita el desarrollo tanto empírico como teórico de la Robótica Educativa.
- Diseñar un prototipo como set educativo de robótica de bajo costo y de alto desempeño.
- Utilizar un entorno de programación de robot que facilite el aprendizaje y desarrollo de programas para personas con pocos conocimientos de Informática.
- Diseñar material didáctico y lúdico como medio de transferencia del conocimiento.

No buscamos crear un curso completo de programación, sino generar un espacio de acercamiento a la tecnología y al mundo de la programación que resulte atractivo para los jóvenes antes de que ingresen a la universidad. Sin embargo, consideramos que esta puede complementarse con las anteriores con la finalidad de reducir la deserción en los primeros años de la carrera.

Líneas de Investigación, Desarrollo e Innovación

- Impacto de las TIC en el proceso de enseñanza-aprendizaje del Nivel Medio.
- Desarrollo de Objetos de Aprendizaje.
- Articulación Universidad - Nivel Medio.
- Robótica Educativa.

Resultados y Objetivos

El desarrollo de esta línea permitirá:

- Fortalecer la relación entre Universidad e instituciones del Nivel Medio.
- Determinar la situación existente respecto a la implementación de las TIC en el aula y contribuir con la incorporación de éstas como medio pedagógico.
- Generar un ambiente de aprendizaje que permita a los alumnos del Nivel Medio integrar distintas áreas del conocimiento, adquiriendo habilidades generales y nociones científicas.
- Constituir un equipo interdisciplinario para la investigación y desarrollo de contenidos educativos y estrategias didácticas en TIC, Robótica y Educación.
- La expansión de las TIC en las Escuelas de Nivel Medio del departamento Chilecito y de la región.
- Disminuir del índice de deserción en el primer año de las carreras incluidas como oferta académica de la UNdeC.

Actualmente se ha diseñado y construido el kit de Robótica Educativa que está compuesto por un módulo principal que aloja al microcontrolador Arduino [4] UNO R3, un sistema de alimentación, un conjunto de sensores

que componen el sistema de percepción del robot y actuadores que permiten modificar las variables del entorno controladas. Se utilizaron sensores para la evasión de obstáculos, seguidores de líneas, detectores de color, de contactos basados en switches, y actuadores como motores o indicadores lumínicos del estado interno del robot.

Este diseño permite que el kit pueda ser armado y desarmado de manera simple, siendo ideal para la experimentación y comprensión de conceptos matemáticos, físicos e informáticos. Para el diseño y construcción se utilizaron piezas de bajo costo y disponibles en el mercado nacional para que en caso de pérdida o daño puedan ser reemplazadas fácilmente.

Se incluye una librería con funciones, programada por el equipo de trabajo, que se anexan al entorno de desarrollo Arduino, con el objeto de brindarle al estudiante un nivel de abstracción superior que le permita programar en un lenguaje más natural e intuitivo que el aportado por el propio entorno.

Se ha definido y especificado el juego que los alumnos deberán resolver y se ha completado el material didáctico interactivo en formato digital que será distribuido durante la realización de los talleres y como material de apoyo y estudio para los estudiantes.

Durante el año 2015, como prueba piloto, se desarrollaron en una Institución de Nivel Medio del Departamento Chilecito, los cuatro talleres planificados. En ese mismo año se adquirió una impresora 3D para el diseño e impresión de los distintos componentes del robot, capacitando sobre su uso a alumnos y docentes que forman parte del proyecto. Se planifica para el año 2016 la construcción de varios kits de robótica educativa con el objeto de replicar los

talleres en otras instituciones del Nivel Medio del Departamento.

Avances de la línea fueron expuestos en el artículo "Robótica educativa: una estrategia para despertar vocaciones tempranas en Informática." aprobado y presentado en el IX Congreso sobre Tecnología en Educación y Educación en Tecnología, Chilecito (La Rioja) 2014; en las III y IV Jornadas Científicas de Estudiantes Investigadores, UNdeC 2014 / 2015 con la participación de un alumno de grado; y en la Sesión de Demos Educativos en el X Congreso de Tecnología en Educación y Educación en Tecnología, Corrientes 2015.

Formación de Recursos Humanos

El equipo de investigación de esta línea de trabajo está compuesto por seis docentes de la Universidad Nacional de Chilecito y dos alumnos avanzados de grado. De los docentes: uno es doctor en ciencias de la computación especializado en cómputo paralelo y tecnología grid, dos docentes que se encuentran desarrollando su tesis de Maestría en Informática uno en el área de Objetos de Aprendizaje y el otro en Mejora del Posicionamiento de Satélites y dos auxiliar docente actualmente egresados de las carreras Ingeniería en Sistemas y Licenciatura en Sistemas de la UNdeC. También participan 2 alumnos avanzados de grado que se encuentran definiendo su trabajo final para la obtención del título de Ingeniero en Sistemas.

Los integrantes son docentes de las asignaturas Algoritmo y Estructuras de Datos, Arquitecturas de Computadoras, Arquitecturas Paralelas, Programación.

Referencias

[1] Josfie Joaquin Brunner and Rocío Ferrada Hurtado, Educación superior en

Iberoamerica 2011, Josfie Joaquin Brunner and Rocío Ferrada Hurtado, Eds. Santiago, Chile: RIL R editores, Oct. 2011.

[2] SPU, "Articulación escuela secundaria educación superior," Blog: Portal de Educación, 2007. [Online]. Available: <http://portales.educacion.gov.ar/spu/cpres/articulacion-escuela-secundaria-educacion-superior/>

[3] Ana Lourdes Acuña, Marfiá Dolores Castro, and Diana Matarrita Obando, "Desarrollo de capacidades para el diseño e implementación de proyectos de robótica educativa en América Latina y el Caribe," Fondo. Reg. para la Inov. Digital en América Latina y el Caribe y la Fund. Omar Dengo, Informe Final de investigación, 2011.

[4] Massimo Banzi, Getting Started with Arduino, 2nd ed. O'Reilly Media, 2011. [Online]. Available: <http://itebooksinfo/book/1338/>

[5] A. H. González and M. C. Madoz, "Utilización de TIC para el desarrollo de actividades colaborativas para la enseñanza de la programación," Jul. 2013. [Online]. Available: <http://hdl.handle.net/10915/27525>

[6] R. Coppo, J. Iparraguirre, G. Feres, G. Ursua, and A. Cavallo, "Sistema didáctico para la enseñanza de la programación con metodologías de aprendizaje basado en problemas," 2011, eje: Tecnología informática aplicada en educación. [Online]. Available: <http://hdl.handle.net/10915/19923>

[7] L. C. De Giusti, F. Leibovich, M. Sanchez, F. Chichizola, M. Naiouf, and A. E.

De Giusti, "Desafíos y herramientas para la enseñanza temprana de concurrencia y paralelismo," Oct. 2013, WIEI - II Workshop de innovación en educación en

informática. [Online]. Available:
<http://hdl.handle.net/10915/31680>

[8] Cristian Rigano and Juan Vivanco, "MHO: un robot de sumo - la primera experiencia en robótica con un LOGO!" Grupo de Robótica y Simulación, Departamento de Ingeniería Eléctrica, UTN-FRBB, Bahía Blanca, Reporte, 2006.

[9] M. A. Junco Rey, R. Swain Oropeza, A. Aceves López, and J. Ramírez Uresti, "RoboCup: el reto tecnológico de monterrey campus estado de méxico," Arequipa, Perú, 2002.

[10] Fernanda B. Carmona, et al. "Robótica educativa: una estrategia para despertar vocaciones tempranas en Informática." IX Congreso sobre Tecnología en Educación & Educación en Tecnología. La Rioja, 2014.

Caracterización del uso de Aulas Virtuales en el Ciclo de Formación Específica de Carreras de Ingeniería

Mirtha Rodríguez de Zar, Cecilia Gallardo, Oscar Quinteros

Facultad de Tecnología y Ciencias Aplicadas, Universidad Nacional de Catamarca

Maximio Victoria 55 – 4700 Catamarca, Argentina
{mirtazar, ceciliagallardo, oequinteros}@tecno.unca.edu.ar

Resumen

El uso de aulas virtuales en entornos de aprendizaje, requiere especialmente del docente las siguientes competencias básicas: pedagógicas, comunicativas y tecnológicas. A su vez, para que la incorporación de las aulas virtuales sea exitosa y que sus resultados constituyan aportes significativos en el proceso de aprendizaje, es una responsabilidad compartida entre técnicos, pedagogos, docentes, estudiantes y la Institución educativa.

Surge así el interés de este equipo de trabajo por conocer cómo es el uso de las aulas virtuales en el Ciclo de Formación Específica de las carreras de ingeniería de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca (UNCA). Se pretende realizar una caracterización, a través de una investigación básicamente descriptiva, a partir del relevamiento de las experiencias que se desarrollan en ese espacio virtual de aprendizaje, de tal manera que posteriormente puedan ser ellas analizadas y evaluadas para medir el impacto sobre el proceso de enseñanza y aprendizaje por competencias que desde el Consejo Federal de Decanos de

Ingeniería de la República Argentina (CONFEDI) se propicia. Los resultados – obtenidos de la propia praxis- pueden ser utilizados para evaluar el nivel alcanzado respecto a estándares en competencias TIC.

Palabras clave: entornos virtuales, competencias docentes, calidad educativa, ingeniería.

Contexto

El presente trabajo de investigación se encuentra enmarcado dentro del Proyecto acreditado Año 2015 por la Secretaría de Ciencia y Tecnología (SECyT) de la UNCA, *Uso de las Aulas Virtuales en el Ciclo de Formación Específica de Carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas - UNCA*. El mencionado proyecto es desarrollado por docentes y alumnos del Departamento de Agrimensura y Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas (Universidad Nacional de Catamarca).

Introducción

En la Sociedad de la Información y el Conocimiento en que nos encontramos inmersos ya no sorprende las constantes - casi diarias- posibilidades de aplicación de las nuevas tecnologías en la vida cotidiana. Internet continúa revolucionando todos los ámbitos de nuestra sociedad y es cada vez más evidente su poderosa influencia en lo educativo.

Especialmente en el nivel universitario, el aula virtual es el lugar donde los educadores y educandos se encuentran para realizar actividades que conducen al aprendizaje en un entorno que debe permitir la interactividad, la comunicación, la interacción, la aplicación de los conocimientos, la evaluación y el manejo de las clases [1].

En la Facultad de Tecnología y Cs. As. se ha puesto a disposición de los docentes (año 2008), una plataforma virtual "open source" con fines educativos denominada Moodle. Es en 2010, a partir del curso de posgrado "Las TIC y las prácticas docentes" dictado en esta Facultad, cuando un grupo de docentes pone de manifiesto la decisión de asumir el reto que su uso implica, al que también se sumaron otros que realizaron su experiencia de manera intuitiva. El Instituto de Informática, dependiente de esta Facultad, es el responsable de la administración y mantenimiento de la plataforma virtual de aprendizaje Moodle.

Hasta febrero del 2015 se registran 47 aulas virtuales correspondientes al Ciclo de Formación Específica de las cuatro carreras de Ingeniería posibles de cursar en dicha Facultad (Ing. en Agrimensura, Ing. Electrónica, Ing. En Informática e Ing. En Minas) y todas ellas en modalidades b-learning. Se observa que aun cuando no se requiere autorización especial, ni presentación de un proyecto o planificación previa, desde hace dos años que la solicitud de activación de aulas

virtuales es muy baja en el Ciclo de Formación Específica de las carreras de Ingeniería.

Esto manifiesta, de alguna manera una tendencia retroactiva en el uso de esta herramienta, lo que no deja de sorprender teniendo en cuenta que los estudiantes universitarios actuales viven en la Sociedad de la información, la cual según reflexiona Marqués [2], *"está modelada por los continuos avances científicos y por la tendencia de la globalización económica y cultural, que cuenta con una difusión masiva de la informática, la telemática y los medios audiovisuales de comunicación en todos los estratos sociales y económicos"*.

El uso de aulas virtuales requiere especialmente del docente competencias básicas: competencias pedagógicas, comunicativas y tecnológicas por la triple función que ellas cumplen: como instrumento facilitador de los procesos de aprendizaje, como canal de comunicación entre formadores y estudiantes y como herramienta para el proceso de la información que permite a los estudiantes aumentar sus competencias digitales.

Actualmente se encuentra trabajando en esta Facultad, un equipo de investigación en el proyecto "Entornos y Herramientas Virtuales para la práctica docente en el Aula de Ingeniería", cuyo principal objetivo es *"promover la incorporación de entornos y herramientas virtuales a la práctica pedagógica de los docentes de la Facultad de Tecnología y Ciencias Aplicadas de la UNCA.[...] La capacitación a docentes, el análisis, estudio y aplicación de las distintas herramientas web2.0 servirán de apoyo a los docentes de las distintas disciplinas para que los mismos puedan adaptarse a los continuos cambios a lo que nos vemos inmersos en la Sociedad de la Información"*

Existen varias experiencias sobre esta temática, por ejemplo, en la Universidad Nacional de Lomas de Zamora y en trabajos de Tesis de carreras de posgrado a nivel nacional, pero se desconoce la realización de investigaciones sobre el desempeño de las aulas virtuales en carreras de Ingeniería de la Facultad de Tecnología y Cs As de la UNCA.

En virtud de ello surge el interés de este equipo de trabajo de conocer cómo es el uso de las aulas virtuales en el Ciclo de Formación Específica de las carreras de Ingeniería de la Facultad de Tecnología y Cs As-UNCA, y efectuar una caracterización a través de una investigación básica descriptiva, con el propósito de que los resultados -obtenidos de la propia praxis- sirvan para la auto-evaluación del proceso enseñanza-aprendizaje y eventualmente de base para la definición de estándares de competencia en TIC. Ellos serán buenos referentes sobre aquello que deben saber y enseñar los docentes, lo que el alumno debe aprender y dominar y lo que la Institución debe favorecer.

Se pretende efectuar un aporte a la gestión de conocimiento, por cuanto, en consonancia con la definición que algunos expertos dan sobre ella, procura identificar, detectar, re-elaborar y poner a disposición de toda la organización los conocimientos y las practicas que aportan valor añadido a los miembros de cualquier entidad. Joyanes [3] dice que la gestión de conocimiento significa entregar a las personas los datos e informaciones necesarias para ser eficientes (eficaces) en su trabajo u organizaciones.

Se podrá así dar un pequeño paso en el difícil campo epistémico de la calidad de la formación académica a través de redes telemáticas.

Líneas de Investigación, Desarrollo e Innovación

El objetivo general del presente trabajo es caracterizar los principales usos de las aulas virtuales, utilizadas como una herramienta de apoyo en el proceso enseñanza-aprendizaje de asignaturas del Ciclo de Formación Específico de las carreras de ingeniera de la Facultad de Tecnología y Cs. As. de la UNCA.

Objetivos Específicos

Turoff [4] afirma que una “clase virtual es un entorno de enseñanza y aprendizaje inserto en un sistema de comunicación mediado por ordenador”. Para que tenga lugar el aprendizaje han de estar presentes ciertos componentes que se definen desde una óptica interdisciplinar [5]: a) funciones pedagógicas (actividades y materiales de aprendizaje, situaciones de enseñanza, apoyo y autorización, evaluación, entre otros); b) las tecnologías apropiadas y la interconexión con el modelo pedagógico; c) la organización social de la comunicación (espacio, calendario y comunidad).

Corresponde, por tanto, atendiendo al objetivo general explicitado anteriormente, definir los siguientes objetivos específicos:

- Identificar las competencias pedagógicas que se perciben en las aulas virtuales.
- Identificar las competencias comunicativas que se evidencian en las aulas virtuales.
- Identificar las competencias tecnológicas que se manifiestan en las aulas virtuales.
- Identificar las estrategias de aprendizaje utilizadas por los alumnos en las aulas virtuales y el

nivel de satisfacción con el cursado virtual.

Resultados y Objetivos

El presente proyecto de investigación se encuentra en una etapa inicial de trabajo, donde se ha realizado un primer relevamiento identificando las cátedras que pertenecen a los ciclos específicos de formación de cada carrera de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas. Luego se ha verificado si cada una de estas cátedras tiene asociada o no, algún aula virtual dentro de la Plataforma Moodle que provee la Facultad.

Se está trabajando en el diseño de las Fichas de Observación y Registro de cada una de las cuatro variables que se identificarán en las aulas virtuales. □

Así también, se ha comenzado con el diseño de una Base de Datos informática, con el objeto de estructurar, almacenar y gestionar los datos recopilados en la presente investigación.

Cabe destacar, que se planea desarrollar una aplicación web que se conectará a la base de datos mencionada y mediante la cual, los integrantes del equipo puedan gestionar directamente los datos recabados, para luego emitir los reportes que sean necesarios.

Formación de Recursos Humanos

El equipo de trabajo se encuentra integrado de la siguiente manera: la dirección está a cargo de una docente con el título de posgrado “Especialista en docencia universitaria de disciplinas tecnológicas”, quien además se encuentra trabajando en su tesis para obtener el título de posgrado “Especialista en entornos virtuales de aprendizaje”. Respecto a los demás miembros del equipo, se cuenta con una Lic. en

Sistemas de Información, docente investigadora de la Carrera de Ingeniería en Informática, con trabajos en el área de desarrollos de sistemas software y especializada en tecnologías web. Otro integrante es Lic. En Sistemas de Información, teniendo a su cargo actualmente la Dirección del Instituto de Informática, donde se administra la plataforma Moodle de esta Facultad. Asimismo conforman el equipo de trabajo, otros docentes investigadores que tienen algunas experiencias académicas con trabajo en aulas virtuales. Por último, se cuenta con la colaboración de tres alumnos, dos de los cuales pertenecen a la carrera de Ingeniería en Informática y otro está cursando la carrera de Ingeniería en Agrimensura.

Referencias y Fuentes

Consultadas

[1] HORTON, W. (2000) Designing web based training Wiley Computer Publisher, New York, NY

[2] MARQUÉS GRAELLS, Pere (2001) "Sociedad de la información. Nueva cultura". Revista Comunicación y Pedagogía, núm. 272, pp. 17-19. Editora: Centro de Comunicación y Pedagogía. Universidad de La Rioja. España.

[3] JOYANES, L. (2003). “Historia de la Sociedad de la Información. Hacia la Sociedad del Conocimiento en Revolución Tecnológica”. Universidad de Alicante. España.

[4] TUROFF, Murray (1995). “Designing a Virtual Classroom”. International Conference on Computer assisted Instruction ICCAT'95. National Chiao

Tung University. Taiwan. Documento electrónico publicado en <https://web.njit.edu/~turoff/Papers/DesigningVirtualClassroom.html>

[5] PULKKINEN, J. et al. (1998) "Pedagogic roles and dynamics in telematics". En Selinger, M. Y Pearson, J. (Eds.) New York, NY

ALVARADO, M. H. (2010) "La integración de las TIC en Instituciones Educativas". Los desafíos de las TIC para el cambio educativo. Metas educativas 2021. Organización de los Estados Iberoamericanos (O.E.I.). Fundación Santillana. Madrid. España.

COLL, C. (2012). "Aprender y Enseñar con las TIC: expectativas, realidad y potencialidades" Metas educativas 2021. Organización de los Estados Iberoamericanos (O.E.I.). Fundación Santillana. Madrid. España.

TORRES-DIAZ, J. et al. (2012). Integración de redes sociales y entornos virtuales de aprendizaje. RED, Revista de Educación a Distancia. Número 35. 1 de enero de 2013. Consultado el [26/02/2016] en <http://www.um.es/ead/red/35/>

Implicancias conceptuales interdisciplinarias en la búsqueda de creatividad y criticidad para ambientes colaborativos y distribuidos soportados por computador y mediados por la gestión del conocimiento.

Isabel Velazquez, Leda Digión, Paola Budán, Analía Aftyka¹, María Lorena Romano Garay¹

Departamento de Informática, FCEYT, UNSE
Av. Belgrano (s) 1912, TE 0385 4509500 Int.1816
{kereyes@unse.edu.ar; ldigion@unse.edu.ar; pbudan@unse.edu.ar}

Resumen

La IHC (Human Computer Interaction) o IPO (Interacción Persona-Ordenador) logra relevancia como disciplina relacionada con el diseño, la evaluación, y la implementación de aplicaciones informáticas para ser usadas por los seres humanos en ambientes colaborativos y distribuidos soportados por computador y la investigación sobre el estudio de los fenómenos más importantes con los que está relacionado.

Si la interfaz es la que facilita la comunicación, la interacción entre los dos sistemas de diferente naturaleza como son, el ser humano y una máquina constituye un campo que genera nuevos interrogantes como ¿Qué y cómo inciden aspectos humanos tales como, los sociales, cognitivos, culturales en la usabilidad de las interfaces? Si asumimos que la propiedad de “usabilidad” exige conocer y entender al usuario y al contexto en el que se desenvuelve superando las diferencias en cuanto a lenguajes diferentes, verbo-icónico en el caso del ser humano y binario en el caso de la computadora, entonces emerge otra preocupación ¿Cómo lograr que las interfaces centradas en el usuario presenten características de usabilidad y

de respuestas efectivas en términos de creatividad y criticidad para potenciar la capacidad de aprendizaje de los seres humanos? ¿Puede la Gestión de Conocimientos constituirse en una herramienta conceptual y técnica capaz de dotar de usabilidad a Ambientes Colaborativos y distribuidos mediados por la Gestión del Conocimiento?

Palabras clave: IHC- Gestión del conocimiento -Ambientes Colaborativos y distribuidos - Criticidad y creatividad

Contexto

La presente propuesta constituye el Proyecto de Investigación del Departamento de Informática de la Facultad de Ciencias Exactas y Tecnologías de la UNSE. El periodo de ejecución se inicia en enero de 2016 y su fecha de finalización se prevé diciembre de 2020. El proyecto continúa la línea de investigación del proyecto denominado Propuesta Metodológica para el desarrollo de Interfaces de Usuario de Sistemas Colaborativos, iniciada en el año 2012 y finalizado en diciembre de 2015 con evaluación final satisfactoria, y

¹ Tesistas de la Maestría en Informática Educativa FCEyT UNSE.

a partir del cual se propone continuar la línea de investigación fortaleciendo la investigación de los fenómenos que surgen de la interacción de CSCW orientados al ámbito educativo entre ellos la creatividad y criticidad. El proyecto estará subvencionado por el Consejo de Investigaciones Científicas y Tecnológicas (CICYT) de la Universidad Nacional de Santiago del Estero (UNSE), Argentina.

Introducción

La exploración y el análisis del estado de la HCI, permite observar el aporte de la Psicología Cognitiva al conceptualizar el aprendizaje como un proceso individual e interno en el que adquiere significatividad el “otro” como mediador entre el sujeto y el conocimiento facilitando el proceso de auto-construcción; el “otro” es la interface que pone en evidencia la capacidad de la persona para captar, procesar, comunicar y generar información. Sin embargo, el avance de la Psicología permite inferir que no sólo la dimensión cognitiva debe estudiarse sino que se debe tender a potenciar el desarrollo humano mediante una perspectiva interdisciplinaria compleja de carácter socio cognitivo y cultural para analizar, comprender y aplicar los conceptos, métodos y técnicas procedentes de otras áreas de conocimiento y contribuir a solucionar los problemas relacionados con el desarrollo de sistemas interactivos orientados a la educación.

En términos técnicos, es el pasaje del Tratado de Card, Moran y Newell de 1983 en el que se crea un modelo analítico con el que evaluar la efectividad de un sistema interactivo a la usabilidad como metodología para crear interfaces más fáciles de usar (Bewley, 1983;

Butler, 1985; Good; Spine; Whiteside; George, 1986) y en la época actual incorporar el concepto de “ubicuidad” con lo que se intensifica la búsqueda de aspectos más subjetivos.

En este proyecto se pretende construir un marco conceptual y referencial interdisciplinar que fundamente la construcción de Interfaces usables en términos de creatividad y criticidad para la gestión del conocimiento en ambientes de trabajo colaborativos y distribuidos con clara intención educativa.

Líneas de Investigación, Desarrollo e Innovación

A partir de la concreción de las actividades previstas en el proyecto se espera obtener los siguientes resultados:

- Alcanzar un mayor entendimiento de los modelos de interacción y la comprensión de la incidencia de los factores humanos, culturales, sociales en el desarrollo de interfaces de sistemas colaborativos.
- Un marco conceptual interdisciplinario que oriente la construcción de interfaces usables y colaborativos para groupware con fines educativos.
- Fortalecer la calidad académica de la comunidad del Departamento de Informática de la FCEyT mediante la investigación de un tema de actualidad y con fuerte tendencia a expandirse en la sociedad del siglo XXI.

Resultados y Objetivos

Los resultados parciales y finales serán socializados y se espera que tengan repercusión en el ámbito académico local y regional por la importancia del tema para el campo de la educación.

Asimismo, las publicaciones en reuniones y revistas científicas sobre el avance de los resultados, serían difundidas en el ámbito local, nacional e internacional para posicionar a la Institución como centro de estudios superiores e investigación.

Entre los principales objetivos que se desean alcanzar son:

- Analizar los efectos sociocognitivos/culturales que se perciben al indagar sobre el estado del arte en el campo de la Ingeniería de la Usabilidad, Gestión del Conocimiento, y del CSCW considerando la complejidad de la interface, la interacción hombre-computadora y la gestión del conocimiento desde una perspectiva interdisciplinaria en busca de vincular la creatividad y la criticidad con ambientes colaborativos y distribuidos.
- Aportar a la reflexión, evaluación y toma de decisiones en el campo de la IHC considerando la incidencia que, los aspectos humanos, sociales, culturales y tecnológicos tienen en las interacciones de los sistemas orientados a la Gestión del Conocimiento y al trabajo colaborativo.
- Identificar los factores humanos, sociales y culturales apropiados para promover la creatividad y criticidad en la generación de interfaces de usuario de sistemas orientados al trabajo colaborativo y distribuido en el ámbito educacional.

Formación de Recursos Humanos

Los recursos humanos que integran el Proyecto pertenecen o están vinculados con la FCEyT de la UNSE. La mayoría de los integrantes son docentes de carreras

de grado y posgrado del Departamento de Informática de la Facultad de Ciencias Exactas y Tecnologías; se incorporan egresados del Profesorado de Informática y de la Licenciatura en Sistema de Información que actualmente se encuentran cursando la carrera de posgrado “Especialización y Maestría en Informática” en calidad de Tesistas. Indicar la estructura del equipo de trabajo de la línea de I/D/I presentada. Indicar el número de Tesis de Posgrado y Tesinas de Grado, en curso y aprobadas en el año anterior, que se relacionan directamente con la línea de I/D/I presentada.

Referencias

- [1] Ausubel; D. 1978. In defense of advance organizers: A reply to the critics. *Review of Educational Research*, 48, 251-257. .
- [2] Bruner, J. 1996. *The culture of education*. Cambridge, MA: Harbard University Press
- [3] Flavell, J. 1979. Metacognition and cognitive monitoring: A new area of cognitivedevelopmental inquiry. *American Psychologist*, 34, 906-911.
- [4] Ford, N. 2008. *Web-Based Learning through Educational Informatics: Information Science Meets Educational Computing*. Information Science Publishing, Hershey. New York. 2008.
- [5] Morin, E. 1998. *Introducción al pensamiento Complejo*. Barcelona. Gedisea.
- [6] Morin, E. 2000. Reforma del pensamiento, transdisciplinarietà, reforma de la universidad. En E. Morin, C. L' Heurreux, A. Paloma & V. Gorr, (Eds.), “Complexus”.
- [7] Pozo, J. 1989. *Teorías cognitivas del aprendizaje*. Morata.

- [8] Sternberg, R. 1998. Metacognition, abilities, and developing expertise: What makes an expert student? *Instructional Science*, 26 (1-2) 127-14
- [9] Penichet, V., Lozano, M., Gallud, J., Tesoriero, R.: *Analysis Models for User Interface Development in Collaborative Systems*. CADUI'08.

Elementos interactivos en los mundos virtuales 3 D

Sattolo Iris¹, Lipera Liliana¹, Romero Juan Carlos¹, Cos Susana¹, Figueroa Sebastián¹, Cisneros Jonatan¹, Altube Alejandro¹, Minutella Darío¹.

Facultad de Informática Ciencias de la Comunicación y Técnicas Especiales
(FICCTE)¹ Universidad de Morón

Cabildo 134, (B1708JPD) Morón, Buenos Aires, Argentina. TE 56272000 Int 189

iris.sattolo@gmail.com, llipera@unimoron.edu.ar, juancarlosromer@gmail.com,
sualcos@gmail.com, sebastianfigue94@hotmail.com.ar, v8javier@gmail.com,
alealtube@hotmail.com, minutelladario@gmail.com

Resumen

Las continuas transformaciones, en las comunicaciones y la tecnología, impactan en los modos de interacción entre las personas, siendo éstas, una de las características que se analizan en el campo de la comunicación. El desarrollo de los mundos virtuales tridimensionales ofrece escenarios para la mejora de las interacciones en la virtualidad, en donde, los usuarios se comunican en tiempo real a través de sus “avatares”. De los avances elaborados por nuestro equipo y de las observaciones realizadas por el público, surgió la inquietud de buscar respuesta a esta pregunta: ¿cómo dar mayor sensación de presencia en los mismos?

Nuestra propuesta es agregar valor estético a los desarrollos, trabajar sobre avatares que reflejen rasgos y patrones de conducta, e incorporar bots conversacionales para aumentar la sensación de presencia e interacción dentro del mundo virtual.

Palabras clave:

Ambientes Virtuales Inmersivos, valor estético, interacción, inteligencia artificial.

Contexto

Este trabajo se presenta como continuación del proyecto de investigación PID 01-002-2014, el cual fuera aprobado por la Secretaria de Ciencia y Tecnología (SECYT) de la Universidad de Morón, llevado a cabo durante los años 2014-2016. A partir de los avances elaborados por el equipo de trabajo y de las observaciones realizadas por el público que participó en las demostraciones plasmadas en el Congreso de Tecnología en Educación y Educación en Tecnología 2015, surgieron inquietudes que nos permitieron plantear nuestra siguiente propuesta: “Ambientes virtuales inmersivos, elementos interactivos.”

Introducción

Se entiende por mundos virtuales inmersivos, también llamados metaversos, a entornos que permiten la recreación de escenarios tridimensionales reales o imaginarios, generados por computadora, con los que el usuario puede interactuar a través de su representación (Avatar) y que produce la sensación de estar adentro. Actualmente, los mundos virtuales en tres dimensiones tienen un importante auge en el ámbito educativo; se presentan como una alternativa que aporta una interfaz más amigable para las poblaciones estudiantiles y académicas, quienes, por lo general, encuentran poco interactivas las plataformas tradicionales de aprendizaje (Moodle y WebCT) debido a que sólo ofrecen entornos bidimensionales.

Estos mundos virtuales están presentes en reconocidas instituciones, entre las cuales se puede nombrar a la Universidad de Barcelona, Universidad de Sevilla y la Universidad a distancia de Madrid. La tendencia, en estas instituciones, de incorporar tecnología a los procesos educativos, no solo se relaciona con el uso de mundos virtuales, sino también de dispositivos móviles y de realidad aumentada. Entre los primeros en incursionar en el tema se puede citar a MIT (Massachusetts Institute of Technology), Universidad Carolina del Sur, Universidad George Mason, Universidad de Nottingham. La Universidad Abierta de Inglaterra, ya desde 1994 hacía sus primeros trabajos en Realidad Virtual y Educación.

La Universidad Autónoma de Madrid a través de su página VirtUAM [1], ofrece los avances y temas desarrollados sobre mundos virtuales en los cuales están trabajando.

Recientemente, La Universidad Politécnica de Madrid suscribió un documento conjunto de participación en el Proyecto de Red UNNOBA-UNPA-UPM sobre el

uso de mundos virtuales en la docencia. [2]

Parte de las experiencias y actividades, dirigidas a este tipo de implementaciones, se apoyan en el New Media Consortium (www.nmc.org), una agremiación de carácter internacional que agrupa a 250 instituciones como: universidades, colegios, museos y organizaciones interesadas en los usos educativos de las tecnologías de información y comunicación y los nuevos medios.

En el Informe horizon2015, “entre las tendencias importantes a mediano plazo”, menciona a los Makerspaces (talleres creativos) como: “la robótica, las aplicaciones de modelado en 3D y las impresoras 3D, han supuesto un cambio de paradigma en cuanto a las habilidades que pueden tener una aplicación y un valor en la vida real. Los talleres en los que se ofrece este tipo de herramientas resultan muy relevantes para el rediseño de los espacios de aprendizaje.” [3]

En nuestra institución, los avances obtenidos hasta el momento fueron desarrollados en Opensim (aplicación servidora multiplataforma y multiusuario), la cual nos permitió construir un espacio virtual para la capacitación en el desarrollo de planes de evacuación. Estos avances fueron reflejados en distintos congresos nacionales.[4] [5] [6] [7]

Este espacio virtual está montado en una isla definida en un servidor de prueba en la Universidad de Morón al cual se accede en forma local a través de una red LAN.

Los mundos tridimensionales, ofrecen una amplia gama de herramientas, las que proporcionan a los usuarios variadas posibilidades de comunicación y conexión, que van desde el chat de texto simple a la interacción colaborativa profesional dentro de los grupos sociales.

La interacción en estos entornos alude a la capacidad del sistema para facilitar el

acceso a los parámetros, los cuales definen la conducta de los objetos y agentes que forman parte del mismo apuntando siempre al diseño de la estructura: usuario-objetos, usuario-usuario y objeto-objeto. No obstante para la reflexión estética en el campo interactivo, diferentes autores manifiestan que, el interés central debe recaer en qué tipo de información y comunicación se está generando y cuáles son los contenidos y la estética de ellas [8].

Esto incluye trabajar en la organización de colores e imágenes y todo aquello que contribuya a incrementar el valor artístico; el uso de imagen, sonido y texto, son elementos esenciales y útiles que pueden lograr una comunicación multisensorial efectiva.

En cuanto al avatar, la representación que el software permite realizar, tiene características básicas. Se puede configurar su vestimenta y rasgos físicos. Puede caminar, correr, pero no pueden realizar movimientos que sean representativos de la persona (saludar con la mano, sonreír, etc.)

Finalmente, estos mundos posibilitan la introducción de “bots”, los cuales representan aplicaciones o programas diseñados para interactuar con otros programas, servicios de internet u operadores humanos, de manera semejante a como lo haría una persona. Se puede programar tareas sencillas como chat o trayectos, hasta tareas complejas (interacción con base de datos y juegos 3D)

El uso de bots puede desarrollar una mayor interacción con las personas, proporcionando ventajas sobre los mundos que sólo utilizan propiedades intrínsecas al objeto, de modo tal que, les otorgaría a estos espacios virtuales mayor dinamismo y credibilidad. La incorporación de bots conversacionales podría crear una primera fuente de resolución de dudas, que sea capaz de dar respuesta a

aquellas preguntas más frecuentes y sencillas a la que pueda enfrentarse un usuario novel. Su uso podría desarrollar una mayor interacción con las personas, a partir de la identificación de sus perfiles humanos y emocionales.

Líneas de investigación, desarrollo e innovación

Desde este proyecto de investigación planteado para los años 2016-2018: “Elementos interactivos en los mundos virtuales 3D” se desprenden diferentes líneas de investigación:

- 1) el uso de técnicas de Diseño gráfico para incrementar la estética en los Mundos virtuales.
- 2) Investigación y aplicación de diferentes métodos que permitan dar más movimiento a los avatares.
- 3) Evaluar el uso de las técnicas de inteligencia artificial y los beneficios que aportarían a los Mundos virtuales inmersivos.
- 4) Plantear con la ayuda de la información obtenida una serie de casos de estudio articulando con cátedras de la facultad de Informática, que permitan analizar los problemas esbozados.

Resultados y objetivos

Con este proyecto se pretende:

Mejorar el ambiente creado durante la investigación anterior para la cátedra de Higiene y Seguridad.

Dejar zonas de capacitación para que los estudiantes participen en el Mundo.

Dejar espacios que favorezcan el desarrollo de las cátedras que se interesen en los metaversos utilizándolos como laboratorios virtuales.

Como resultado de las diferentes etapas se elaborará material para participar en distintos congresos tales como WICC (Workshop de Investigadores en Ciencias de Computación), CACIC (Congreso

Argentino de Ciencias de la Computación), TE&ET (Tecnología en educación y Educación en Tecnología) como así también en otras cortes, en las cuales el tema sea de interés.

Promover la construcción y dictados de cursos de formación o asesoramiento.

Formación de Recursos Humanos

Este proyecto se encuentra financiado por la Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales de la Universidad de Morón. A su vez propicia la formación de recursos, con la participación de estudiantes de grado para la continuación de las líneas de investigación relacionadas. Este PID está compuesto por cuatro investigadores y docentes de dicha facultad y dos tesis de las carreras de Informática. A su vez están trabajando en esta línea, estudiantes a los últimos años de las carreras de sistemas.

Referencias

[1] <http://aida.ii.uam.es/virtuam/>

[2] <http://serviciosgate.upm.es/laboratoriovirtuales/es/content/participaci%C3%B3n-en-el-proyecto-de-red-unnoba-unpa-upm>

[3] <http://cent.uji.es/octeto/node/4469>

[4] Sattolo Iris, y otros, *Primeros pasos en el desarrollo de ambientes virtuales inmersivos de aprendizaje utilizando software libre*. ISBN 978-987-23963-1-2 P 566-573.

[5] Sattolo et al, *Modelo de análisis de un ambiente virtual de aprendizaje*

inmersivo para el desarrollo de planes de evacuación. Disponible en:

<http://www.teyet2014.undec.edu.ar/Libro-de-ActasTEYET2014.pdf>

[6] Altube et al 2015. *Desarrollo de planes de evacuación utilizando un ambiente virtual inmersivo interactivo*.

Disponible en:

<http://teyet2015.unne.edu.ar/files/TEyET2015.pdf>

[7] Lipera y otros, *Ambiente Virtual inmersivo interactivo* Disponible en:

Libro de actas Jatic 2015 ISBN 978-987-23963-2-9

http://jatic2015.ucaecemdp.edu.ar/libro_actas.php

[8] Claudia Giannetti Libro *Estética, Ciencia y Tecnología Pontificia* Universidad Javeriana 2005

Herramientas informáticas orientadas a la enseñanza y el aprendizaje.

Cecilia De Vito, Germán Osella Massa, Laura Yamel, Paula Lencian, Tamara Ahmad, Mónica Sarobe, Trinidad Picco, Eliana Serrano, Claudia Banchoff

Instituto de Investigación y Transferencia en Tecnología (ITT), Escuela de Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)

Sarmiento y Newbery (CP 6000), Junín, Buenos Aires, Argentina. Teléfonos (0236) 4636945/44
{cecilia.devito,german.osella, lyamel, paula.lencina }@itt.unnoba.edu.ar
{tamaraahmad, monicasarobe, trinidadpicco, elianaserrano, claudiab}@unnoba.edu.ar

Resumen

El amplio abanico de posibilidades tecnológicas disponible en la actualidad no siempre es aprovechado al máximo en el contexto educativo. En esta línea de investigación se propone explotar nuevas formas de aplicar los recursos disponibles para facilitar tanto al docente la transmisión del conocimiento así como al alumno poder adquirirlo, buscando desarrollar nuevas herramientas informáticas que simplifiquen el armado del material usado tanto dentro de sus clases así como de referencia para su posterior estudio fuera del aula. También se proponen herramientas orientadas a aportar dinamismo a las explicaciones del docente, aprovechando la disponibilidad de computadoras cada vez más potentes que permiten la obtención de resultados casi instantáneos al realizar tareas complejas, buscando así clases que sean interactivas, donde la participación de los alumnos adquiera un rol protagónico y no solo sean receptores en una comunicación unidireccional.

Palabras clave:

Educación. Herramientas informáticas. Innovación.

Contexto

Esta línea de investigación forma parte del proyecto “Herramientas didácticas y nuevas formas de interacción para la inclusión de TICs en la enseñanza” aprobado por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA en el marco de la convocatoria a Subsidios de Investigación Bianuales (SIB2015).

El proyecto se desarrolla en el Instituto de Investigación en Tecnologías y Transferencia (ITT) dependiente de la mencionada Secretaría, y se trabaja en conjunto con la Escuela de Tecnología de la UNNOBA.

El equipo está constituido por docentes e investigadores pertenecientes al departamento de Informática y Tecnología y estudiantes de las carreras de Licenciatura en Sistemas, Ingeniería en Informática y Licenciatura en Diseño Gráfico de la UNNOBA.

Introducción

El rápido avance de la tecnología junto con la indudable integración de la misma

a la vida cotidiana contrasta con la inercia que se percibe en ciertas instituciones, donde la manera histórica en que se hacen las cosas se antepone fuertemente a los cambios innovadores que la modernidad permite. Puntualmente, en el ámbito educativo se observa un fuerte arraigo a la forma en que tradicionalmente se arma el material de estudio usado tanto en las explicaciones dentro del aula así como para su posterior análisis fuera de la institución. Desde los dibujos realizados sobre láminas de cartulina, pasando por los retroproyectores y sus transparencias de acetato, las videocaseteras con sus cassettes VHS, llegando a los modernos cañones con sus presentaciones al estilo PowerPoint, la tecnología ha ido perfeccionando en esencia la misma forma de comunicación. Eso sí, con los avances se facilitó la edición y, sobre todo, la duplicación y transmisión del material elaborado. Más allá del medio, sin embargo, la comunicación con el material presentado ha sido siempre de carácter unidireccional, no permitiendo interactuar con el mismo.

En Informática, por ser una ciencia apoyada en el uso de máquinas capaces de realizar cálculos, es común que surja la necesidad de comprender la ejecución de un fragmento de un programa, el comportamiento temporal de determinada estructura de datos o la interacción entre distintas partes de un sistema observando como dicha máquina actúa en cada uno de esos casos. Resulta invaluable contar con la posibilidad de interactuar con el material que se utilice para presentar estos temas, pudiendo así observar algunos comportamientos complejos que de otra forma se perderían en un medio estático [1]. El uso de herramientas como Microsoft PowerPoint [2], LibreOffice Impress [3], Apple Keynote [4], Prezi [5] o similares, pensadas para la creación y

diseño de presentaciones visualmente atractivas, no resulta simple para el armado del material didáctico, sobre todo si lo que se desea mostrar es la ejecución de un programa de computadoras junto con sus variables, instrucciones, los resultados que produce, etc. [6]. El código escrito en una diapositiva no puede ser ejecutado o modificado para observar como eso altera su comportamiento. Carece de interacción y el docente deberá prever de antemano todas las posibilidades que los alumnos le puedan plantear.

En las carreras de informática se espera que en las materias iniciales los alumnos desarrollen la capacidad para visualizar en sus mentes la ejecución de un programa [7]. Para esto, resulta conveniente ayudarlos con el uso de diagramas y animaciones que muestren la secuencia de instrucciones seguidas por un programa y que resalten los resultados obtenidos [8]. Si lo que se quiere es mostrar la manipulación de una estructura de datos, puede resultar conveniente escribir dicha visualización mediante sentencias en un lenguaje de programación en contrapartida a crearla manualmente usando el ratón, apuntando, haciendo clic y arrastrando dibujos, pues ya se cuenta con los propios algoritmos de dicha estructura, que no son otros que los que se quiere ilustrar [9].

Por otro lado, existe la realidad de que los docentes de una asignatura suelen producir contenidos que serán utilizados tanto dentro del aula así como material de estudio para sus alumnos. Es típico que preparen sus clases apoyándose en presentaciones construidas con ayuda de de las herramientas ya mencionadas, permitiendo incorporar con facilidad contenidos multimediales para clarificar las ideas complejas que intentan transmitir. Si el esfuerzo para preparar el

material a usar durante la clase es grande, probablemente sea el único material disponible para los alumnos. En consecuencia, se terminan creando presentaciones sobrecargadas de texto para que sean auto-contenidas e incluso puedan ser comprendidas por alguien que no estuvo presente durante la explicación. Esto termina entorpeciendo o quitando dinamismo a las clases [10]. Por lo contrario, presentaciones pensadas para ser usadas únicamente durante una explicación probablemente no sirvan de nada sin el docente explicándolas.

En un mundo ideal, deberían existir dos tipos de materiales: Por un lado, el que el docente usa sólo durante las ponencias, con bajo contenido de texto pero abundancia de gráficos, código y diagramas, con el que éste complementa su oratoria. Por otro lado, apuntes en forma de libro, que incorpore el contenido anterior completándolo con la prosa necesaria para que el alumno lo pueda comprender y asimilar. Esto conlleva varias dificultades: No sólo se deberá invertir el doble del esfuerzo en preparar ambos tipos de materiales sino que deberá tenerse sumo cuidado de mantener el contenido sincronizado, pues cada cambio realizado en el material de un tipo deberá reflejarse en el del otro tipo. Tristemente, no tenemos constancia de la existencia de herramientas informáticas desarrolladas para tal fin, que a partir de un único documento que sirva como fuente de verdad permitan generar tanto el material de estudio como presentaciones que sean visualmente atractivas para acompañar el dictado de clases.

Líneas de Investigación, Desarrollo e Innovación

Las líneas de investigación, desarrollo e innovación que se proponen en este trabajo están centradas principalmente en el estudio y desarrollo de nuevas herramientas informáticas enfocadas en apoyar al docente en la enseñanza y transmisión del conocimiento. Puntualmente, se enfoca la atención en desarrollos orientados a la enseñanza de los conceptos asociados a la Informática a partir de la aplicación de soluciones innovadoras. Se estudiará la factibilidad para extender los resultados obtenidos para ser aplicados a la enseñanza de otras disciplinas.

Resultados y Objetivos

Se ha desarrollado una herramienta denominada AutoPython que muestra un símil del shell de Python en el que se ha automatizado el ingreso de fragmentos de código previamente preparados, simulando el comportamiento de un operador humano escribiéndolos, haciendo resaltado de sintaxis en colores, pudiendo repetir sentencias previamente ingresadas e incluso cambiar a modo interactivo, permitiendo durante una presentación el ingreso por teclado de nuevas sentencias no previstas. Esta herramienta permite la demostración de ejemplos escritos en el lenguaje Python, ejecutándolos rápidamente y sin pausas o interrupciones por culpa de errores de tipeo, apuntando al dictado de clases en donde se interactúa gran parte del tiempo con el intérprete real para mostrar resultados.

Se está diseñando una herramienta para la creación de material que idealmente permitan volcar los contenidos de una lección en un único documento para luego generar a partir de éste tanto el material de estudio para ser consumido como un capítulo de un libro

así como para extraer ciertos fragmentos del mismo y armar con ellos una presentación para ser empleada durante una clase. Esta herramienta permitiría generar material visualizable tanto como un documento estático (en formato PDF, por ejemplo) o uno interactivo (creando un documento HTML5 conteniendo código JavaScript).

Creación de visualizaciones orientadas primariamente a estudiar y comprender el funcionamiento de varias estructuras de datos, mostrando distintos niveles de abstracción y con diferentes lenguajes de programación. Aún no se ha comenzado a trabajar en esta línea.

Formación de Recursos Humanos

Esta línea de investigación cuenta con 6 docentes investigadores y 3 Alumnos avanzados de las carreras Licenciatura en Sistemas e Ingeniería en Informática, los 3 con becas de PROMINF (Convocatoria realizada por la Secretaria de Investigación y Transferencia UNNOBA) relacionados con el proyecto de investigación en el cual se enmarca la línea. Se encuentra en desarrollo una Tesina de grado.

Referencias

- [1] Thomas L. Naps et al., "Exploring the role of visualization and engagement in computer science education". Working group reports from ITiCSE on Innovation and technology in computer science education. ITiCSE-WGR '02. Aarhus, Denmark: ACM, 2002, pp. 131-152.
- [2] Microsoft PowerPoint. Sitio web: <http://office.microsoft.com/es-es-ar/powerpoint/>
- [3] LibreOffice Impress. Sitio web: <https://es.libreoffice.org/descubre/impress/>
- [4] Apple Keynote. Sitio web: <https://www.apple.com/es/mac/keynote/>
- [5] Prezi. Sitio web: <http://prezi.com/>
- [6] Miller, B. N., & Ranum, D. L. (2012, July). Beyond PDF and ePub: toward an interactive textbook. In Proceedings of the 17th ACM annual conference on Innovation and technology in computer science education (pp. 150-155). ACM.
- [7] Vihavainen, A., Helminen, J., & Ihantola, P. (2014, November). "How novices tackle their first lines of code in an ide: Analysis of programming session traces". In Proceedings of the 14th Koli Calling International Conference on Computing Education Research (pp. 109-116). ACM.
- [8] Guo, P. J. (2013, March). "Online python tutor: embeddable web-based program visualization for CS education". In Proceeding of the 44th ACM technical symposium on Computer science education (pp. 579-584). ACM.
- [9] Carmer, B., & Rosulek, M. (2015, October). "Vamonos: Embeddable visualizations of advanced algorithms". In Frontiers in Education Conference (FIE), 2015. 32614 2015. IEEE (pp. 1-8). IEEE.

- [10] Penciner, R. (2013). “Does PowerPoint enhance learning?”. CJEM, 15(02), 109-112.

Prácticas de m-learning desde un enfoque ecológico

Marta C. Fennema^{1,2}, Susana I. Herrera², María I. Morales², Jorge L. Goñi², Rosa A. Palavecino², Ivana Irurzun², Alvaro J. Carranza²

(1) *Universidad del Norte Santo Tomás de Aquino*
crisfen@yahoo.com

(2) *Instituto de Investigación en Informática y Sistemas de Información, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero*
mines_morales@yahoo.com.ar, {sherrera, rosypgg}@unse.edu.ar, jlgoni@live.com.ar, ivanairurzun@gmail.com, carranza1903@hotmail.com

Resumen

En este artículo se presentan las acciones y resultados obtenidos de investigar sobre cuáles son los principios y tecnologías de la Computación móvil apropiados aplicar en el diseño de experiencias de aprendizaje.

Los principales objetivos consisten en: a) diseñar experiencias de aprendizaje basadas en dispositivos móviles para los niveles educativos medio y superior, b) desarrollar aplicaciones innovadoras colaborativas para m-learning.

La investigación se inició en el año 2012 y se encuentra en su etapa final. Se ha desarrollado un marco para el análisis, diseño y evaluación de estrategias de mobile-learning (m-learning) denominado MADE-mlearn. Se ha incluido una aplicación, llamada MADE-mlearn App que permite realizar la evaluación de experiencias en forma automática.

Se diseñaron y se llevaron a cabo experiencias de aprendizaje con dispositivos móviles, de distintos modos de interacción y grados de complejidad para cursos de secundaria de zonas rurales y en cursos de grado y posgrado universitarios. Éstas fueron diseñadas usando el MADE-mlearn.

Se ha desarrollado la aplicación Educ-Mobile, juego educativo móvil colaborativo, apropiada para experiencias en cursos con reducido número de alumnos.

En la última etapa de esta investigación, se diseñarán e implementarán experiencias para el área de Matemática (nivel universitario) y de Programación (nivel medio y universitario).

Palabras clave: m-learning, computación móvil, MADE-mlearn, Educ-Mobile.

Esta investigación se lleva a cabo en el marco del proyecto denominado “Optimización de la calidad de los Sistemas Móviles mediante la implementación de nuevas arquitecturas, realidad aumentada, técnicas de visualización y redes móviles Ad-Hoc. Aplicaciones en m-learning y en gestión del conocimiento”. A su vez, el proyecto forma parte de un Programa de Investigación en el área Informática cuyo objetivo consiste en realizar investigación aplicada sobre nuevas tecnologías que permitan mejorar la calidad tanto del proceso como del producto software.

El equipo de investigación del proyecto marco está conformado por docentes de universidades del NOA: del Instituto de Investigaciones en Informática (IISI) de la Facultad de Ciencias Exactas de la Universidad Nacional de Santiago del Estero (UNSE), del Departamento de Computación de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta (UNSa) y de la Universidad del Norte Santo Tomás de Aquino (UNSTA).

En esta línea de investigación específica referida a m-learning participan investigadores del IISI y de la UNSTA. Cuenta con asesoramiento de investigadores del Instituto de Investigación en Informática LIDI de Facultad de Informática de la Universidad Nacional de La Plata, en temas referidos al aprendizaje mediado por tecnologías. También cuenta con asesoramiento de una investigadora de la Universidad Nacional de Mar del Plata, en desarrollo de objetos de aprendizaje.

Está financiado por el Consejo de Ciencia y Técnica de la Universidad Nacional de Santiago del Estero, durante el período Enero 2012-Diciembre 2015.

1 Contexto

2 Introducción

En la sociedad actual, de constantes avances tecnológicos, las personas necesitan de un aprendizaje continuo que les permita desempeñarse de manera eficaz tanto en el ámbito profesional como en la vida cotidiana; es decir una sociedad del conocimiento que demanda de las personas gran capacidad de aprendizaje, adaptabilidad y flexibilidad. La educación ha experimentado grandes cambios en las últimas décadas debido a las nuevas formas de acceso y difusión de la información y comunicación y en los últimos años, las tecnologías móviles han modificado aún más el panorama educativo, dotándolo de movilidad, conectividad, ubicuidad y permanencia.

A nivel mundial se han llevado a cabo experiencias de aprendizaje basado en dispositivos móviles que reportan resultados positivos, pero en general sólo se presentan un conjunto de herramientas y estrategias en tanto las teorías que lo sustentan aún continúan en estudio. Varios grupos de investigación trabajan sobre el tema y han resultado de antecedente para este proyecto [1, 2, 10, 11, 23, 24, 28, 29, 36, 37], en particular se toman las fundamentaciones de m-learning basadas en conceptos de ecología socio-cultural [29, 30] y en el aprendizaje colaborativo. En la Argentina continúan siendo escasas las experiencias de m-learning realizadas, siendo propuestas aisladas basadas en casos concretos [3, 4, 5, 6, 26, 34].

En los inicios de esta investigación, el estudio estuvo centrado en la educación superior de grado y principalmente de posgrado, donde los dispositivos móviles resultan de gran utilidad para los estudiantes que, por su trabajo, poseen limitaciones de tiempo para dedicar al estudio. [12, 13, 15, 38, 40]. Luego se extendió a otros niveles educativos.

En el primer año de esta investigación, se desarrolló una herramienta, denominada MADE-mlearn, que sirve de marco o ecosistema para el análisis, diseño y evaluación de experiencias de m-learning, [17, 21]. Ésta abarca no sólo aspectos tecnológicos sino también socio-culturales, pedagógicos y de interacción. Plantea tener en cuenta características y subcaracterísticas que definen los aspectos antes mencionados, agrupándolos en Categorías.

Además se estudiaron los diferentes modos de interacción que se presentan en las experiencias de m-learning. Según Woodill, ellas pueden ser de tres modos diferentes [40]: recuperación de información (Modo 1), recopilación y análisis de información (Modo 2) y comunicación, interacción y colaboración en redes (Modo 3),

En [12, 13, 15] se presentó la implementación de prácticas de m-learning modo 1, modo

recuperación de información, en el curso de posgrado Enseñanza de la Tecnología de la carrera Especialización en Enseñanza en Tecnologías de la FCEyT de la UNSE. En este caso, al realizar el estudio del ecosistema se determinó que prevalecían los sistemas operativos Android y se diseñaron las prácticas orientadas a dispositivos que corren bajo este sistema operativo. Como herramientas se usaron alimentadores de información o feeders para móviles (Google Reader), lector de e-books para móviles (Libros de Google, Aldico), bibliotecas o libraries para móviles (IEEE mobile).

En [17, 33] se presentaron los resultados alcanzados de la puesta en práctica de estrategias de m-learning del Modo 1 para la enseñanza de la Matemática en el nivel medio, en una zona rural con escasos recursos tecnológicos. Se trabajó con teléfonos móviles sencillos (Nokia C2 y Samsung Chat), todos con Bluetooth y los recursos involucrados fueron gif animados el desarrollo de los temas: Rectas paralelas y perpendiculares y Clasificación de ángulos.

En [17, 19, 20] se expusieron los resultados obtenidos del desarrollo de experiencias de modo 3, Comunidad de Aprendizaje, que consiste en el uso de Educ-Mobile [18], una aplicación creada en el grupo de desarrollo del mencionado proyecto de investigación, y que consiste en juego móvil interactivo y colaborativo en línea, basado en posicionamiento. Esta experiencia fue llevada a cabo en cursos sobre la enseñanza de la Tecnología, tanto en una carrera de posgrado como en una carrera de grado, para afianzar conocimientos (repaso de los temas tratados en clase) como para evaluación de los mismos; fomentando el trabajo colaborativo entre pares.

En [28] se describen experiencias de m-learning en Modo 2 que se llevaron a cabo en asignaturas de Matemática de las carreras de Ingeniería y Licenciatura en Sistemas de Información de la UNSE. Dichas experiencias fueron diseñadas para su aplicación en cursos numerosos y abarcan un conjunto de actividades para la comprensión y autoevaluación de contenidos.

En el año 2015, con objeto de obtener resultados más sólidos se repitieron prácticas de m-learning colaborativo y se avanzó en el diseño de una herramienta que permitiera evaluar experiencias de m-learning con MADE-mlearn. Se diseñó la aplicación MADE-mlearn, la cual está siendo validada en la actualidad. Se comenzó el estudio de la relación entre objetos de aprendizaje [9] y recursos móviles para el aprendizaje.

Además, se inició la integración de conocimientos con el proyecto paralelo del programa de investigación, referido a gestión del conocimiento. En dicho sentido, se avanzó sobre

el diseño de aplicaciones móviles que colaboran con la gestión del conocimiento forestal, trabajando con el Instituto de Investigación de la Llanura Chaqueña de la UNSE [16].

En el año 2015 se diseñaron prácticas de m-learning para asignaturas que pertenecen a las diversas áreas curriculares de carreras de Ingeniería; las mismas fueron diseñadas usando MADE-mlearn. Algunas de ellas ya fueron implementadas y otras se implementarán en el corriente año [22].

3 Líneas de investigación y desarrollo

La principal línea que guía esta investigación radica en el análisis y diseño de estrategias de m-learning. Esta línea implica:

- Diseñar un marco para el análisis, diseño y evaluación de experiencias de m-learning, desde un enfoque ecológico (considerando tanto aspectos tecnológicos como pedagógicos y socio-culturales).
- Construir una herramienta que permita evaluar las experiencias de m-learning, basada en el marco, en forma automatizada.
- Diseñar experiencias de m-learning que sean innovadoras, que beneficien ampliamente el proceso de aprendizaje desde los resultados de aprendizaje y la motivación (modo 1 a 3).
- Desarrollar aplicaciones móviles y objetos de aprendizaje móviles que apoyen a las experiencias de m-learning.

4 Objetivos y resultados

El objetivo general de esta investigación es:

Proponer un marco de análisis, diseño y evaluación de experiencias de m-learning que contenga:

- *aspectos claves para clasificar, comparar, estudiar, elaborar y evaluar propuestas educativas de m-learning, desde un amplio enfoque que contemple cuestiones pedagógicas, tecnológicas y socio-culturales; y*
- *aspectos claves para desarrollar aplicaciones de m-learning.*

Los resultados que se obtuvieron en los cuatro primeros años de desarrollo del proyecto (2012-2015) son:

- Ecosistema móvil del NOA [12, 13, 14, 15].
- Diseño e implementación de estrategias de m-learning, modos 1 y 2 en el área Matemática [17, 28, 33].

- Marco para el Análisis, Diseño y Evaluación de experiencias de m-learning, MADE-mlearn [14, 17, 21].
- Diseño e implementación de estrategias de m-learning de modo 3 (colaborativas), usando el MADE-mlearn [17, 19, 20].
- Desarrollo de la aplicación Educ-Mobile, juego educativo y colaborativo para m-learning [17, 18, 19, 20].
- Diseño preliminar de aplicaciones móviles que contribuyen a la Gestión del Conocimiento en instituciones de investigación rural
- Diseño e implementación de estrategias de m-learning de modo 1, 2y3 para carreras de Ingeniería en diferentes áreas curriculares [22], usando MADE-mlearn. En Tabla 1 se muestran áreas curriculares, carreras y asignaturas para realizar las prácticas.

TABLA 1. PRÁCTICAS DE M-LEARNING PARA INGENIERÍA.

Área Curricular Ingeniería	Carrera de Grado / Universidad	Asignatura
Ciencias Básicas	<ul style="list-style-type: none"> • Ingeniería Eléctrica • Ingeniería Electromecánica • Ingeniería Electrónica • Ingeniería Civil • Ingeniería en Agrimensura • Ingeniería Hidráulica • Ingeniería Industrial • Ingeniería Vial UNSE	<ul style="list-style-type: none"> • Álgebra y Geometría Analítica • Álgebra Lineal
Tecnologías Básicas	<ul style="list-style-type: none"> • Ingeniería en Informática UNCA	<ul style="list-style-type: none"> • Estructura de Datos
Tecnologías Aplicadas	<ul style="list-style-type: none"> • Ingeniería en Agrimensura UNSE	<ul style="list-style-type: none"> • Mediciones Especiales
Complementarias	<ul style="list-style-type: none"> • Ingeniería en Informática UCSE	<ul style="list-style-type: none"> • Metodología de la Investigación
	<ul style="list-style-type: none"> • Ingeniería Civil UNSE	<ul style="list-style-type: none"> • Inglés

Es importante destacar que todos estos resultados parciales permitieron la participación en eventos de relevada importancia como ser:

- The 18th International Conference on Interactive Collaborative Learning (ICL2015), marco 2015 World Engineering Education Forum (WEEF2015) [19].
- International Conference on Collaboration Technologies and Systems (CTS 2014) [20].
- Presentación del libro “*Aprendizaje basado en dispositivos móviles. Experiencias en la Universidad Nacional de Santiago del Estero.*” en la V Feria Provincial del Libro de la Provincia de Santiago del Estero en Octubre de 2014 [17].
- Conferencias para los docentes en la Feria de las Ciencias de la Provincia de Santiago del Estero, años 2013 y 2014.

En este último año del proyecto se desarrollarán las actividades siguientes usando MADE-mlearn:

- Se continuará con la implementación de experiencias de m-learning para las diversas áreas de las carreras de ingeniería, en el nivel universitario y medio.
- Se llevarán a cabo experiencias de m-learning para Programación para carreras de grado que serán abordadas en colaboración con el III-LIDI de la UNLP.
- Se diseñarán experiencias de m-learning que involucren objetos de aprendizaje y se desarrollarán aplicaciones móviles necesarias para ello.
- Se continuará con el diseño de recursos de aprendizaje para ser utilizados en zonas de recursos escasos.
- Se diseñarán aplicaciones móviles para la Gestión del Conocimiento Forestal.
- Se estudiarán y analizarán los resultados de aprendizaje y la motivación.

5 Formación de recursos humanos

La Directora del proyecto pertenece a la UNSTA. La Codirectora, al Departamento de Informática de la UNSE. Las asesoras en Tecnologías Informáticas aplicadas a la Educación pertenecen a la UNLP y a la UNMdP. El resto de los integrantes son docentes investigadores de la Universidad Nacional de Santiago del Estero. También están incorporados al proyecto profesionales informáticos del medio y alumnos avanzados de carreras de Informática que brindan apoyo a la implementación de las experiencias de m-learning y que reciben apoyo en su inicio en tareas de investigación.

Esta investigación contribuye al desarrollo de dos tesis doctorales en Ciencias Informáticas (UNLP), dirigidas/codirigidas por las asesoras y por la directora del proyecto. Asimismo se realiza un trabajo final de Especialización en Enseñanza de las Ciencias Exactas (UNSE). También se preparan trabajos finales de alumnos de la Licenciatura en Sistemas de Información y de la Maestría en Informática Educativa; todas carreras de la FECEyT-UNSE.

6 Referencias

1. Bannan, B., Peters, E., Martinez, P. *Mobile, Inquiry-based learning and geological observation: An exploratory study*. International Journal of Mobile and Blended Learning, 2(3), 13-29. 2010.
2. Blythe, M., Reid, J., Wright, P., Geelhoed, E. *Interdisciplinary criticism: analysis de experience of Riot! A location-sensitive*

digital narrative. Behaviour and Information Technology, 25, 2, 127-139, 2006.

3. Cataldi, Z & Lage, F. *TICs en Educación: Nuevas herramientas y nuevos paradigmas. Entornos de Aprendizaje Personalizados en dispositivos móviles*. Congreso TE&ET. ISBN 978-987-28186-0-9. Pergamino, Buenos Aires, 2012.
4. Cukierman y Otros. *Informe Final del Proyecto "Integración de la Tecnología Móvil a los Entornos Virtuales de Enseñanza y de Aprendizaje"*. Comisión de Investigaciones Científicas del Gobierno de la Pvcia. de Buenos Aires. Bs. As., 2008.
5. Cukierman, U. y Otros. *Una experiencia de uso de celulares en un curso de articulación escuela media y universidad en modalidad a distancia*. VirtualEduca, Brasil, 2007.
6. Cukierman, U. y Virgili, J. *La Tecnología educativa al servicio de la educación tecnológica*. UTN, Buenos Aires, 2010.
7. Ferreira Szpiniak, A., Sanz, C. *Hacia un modelo de evaluación de entornos virtuales de enseñanza y aprendizaje. La importancia de la usabilidad*. TE&ET, Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología. 2010.
8. Fox, S. *Mobile Learning: Is Your Business Ready for It?* Technology Evaluation Centers. Disponible en: <<http://www.technologyevaluation.com/research/articles/mobile-learning-is-your-business-ready-for-it-21858/>>, Dic. 2010, consulta: 05/01/2011.
9. Goñi, J., Herrera, S., Ferreira, R. *Objetos de Aprendizaje en función de su abordaje pedagógico*. Asociación Regional de Economía y Sociedad del Noroeste Argentino (ARESNOA). Salta, 2014.
10. Gwee, S., Chee, Y. S., Tan, E. M. *The Role of Gender in Mobile Game-Based Learning*. International Journal of Mobile and Blended Learning, 3(4), 19-37. 2011.
11. Hansen, F., Kortbek, K., Gronbaek, K.. *Mobile Urban Drama for Multimedia-Based Out-of-School Learning*. ACM. 978-1-4503-0424-5, 2010.
12. Herrera, S. I. & Fennema, M. C. *Tecnologías Móviles Aplicadas a la Educación Superior*. Congreso Argentino de Ciencias de la Computación. La Plata, 20.
13. Herrera, S. I., Fennema, M. C. & Sanz, C. V. *Estrategias de m-learning para la formación de posgrado*. Congreso TE&ET. ISBN 978-987-28186-0-9. Pergamino, Bs. As., 2012.
14. Herrera, S. I., Fennema, M. C., Carrizo, M. I., Ponce, G. *Computación móvil aplicada al aprendizaje. Un marco de análisis del m-learning*. VIII Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA. Tucumán, Septiembre 2012.

15. Herrera, S. I., Goñi, J. L. & Fennema, M. C. *El m-learning en la educación universitaria de posgrado*. Jornadas de Ingeniería del NOA. Catamarca, 2011.
16. Herrera, S. I., López, G., Ríos, M. *Nuevas Tecnologías Móviles para la Gestión del Conocimiento*. Primer Congreso Internacional Gran Chaco Americano (CONGRACHA). Santiago del Estero. 2014.
17. Herrera, S. I., Morales, M. I., Sanz, C. & Fennema C. *Aprendizaje basado en dispositivos móviles. Experiencias en la Universidad Nacional de Santiago del Estero*. Ed. EDUNSE. ISBN 978-987-1676-18-7. Santiago del Estero. 2014.
18. Herrera, S. I., Najar, P., Morales, M.I., Sanz, C., Fennema, M. C. *Educ-Mobile. Juego educativo colaborativo para m-learning. DEMO*. Congreso TE&ET. ISBN 978-987-28186-0-9. Chilecito, La Rioja, 2014.
19. Herrera, S. I., Sanz, C. *Collaborative m-learning practice using Educ-Mobile*. International Conference on Collaboration Technologies and Systems (CTS). Ed. IEEE. Pp 363 – 370. ISBN: 978-1-4799-5157-4. Minneapolis, MN, USA. 2014
20. Herrera, S. I., Sanz, C. *Práctica de m-learning colaborativo usando Educ-Mobile CONAIISI (CONFEDI)*. San Luis. 2014.
21. Herrera, S., Fennema, C., Sanz, C. *MADE-mlearn: un marco para el análisis, diseño y evaluación de experiencias de m-learning en el nivel de postgrado*. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología. N° 10. ISSN 1850-9959. La Plata, 2013.
22. Herrera, S. Fennema, C. Morales, M. María Inés, Palavecino, R. Goldar, E. *Mobile Technologies in Engineering Education*. The 18th International Conference on Interactive Collaborative Learning (ICL2015). Florencia, Italia, 2015.
23. Herrington, A. and Herrington, J. *Authentic mobile learning in higher education*. In: AARE 2007. International Educational Research Conference, Fremantle, Western Australia, 2007.
24. IMS Global Learning Consortium. *Instructional Management Systems*. Disponible en : <http://www.imsproject.org/specifications.HTML>
25. Kalloo, V., Mohan, P. *An Investigation Into Mobile Learning for High School Mathematics*. International Journal of Mobile and Blended Learning, 3(3), 59-76. 2011.
26. Lliteras, A., Challiol, C., Gordillo S. *Juegos Educativos Móviles Basados en Posicionamiento: Una Guía para su Conceptualización*. 13th Argentine Symposium on Software Engineering, 41 JAIIO, La Plata, 2012.
27. Morales, M. I., Herrera S. I., Fennema, M. C., Goñi, J. L. *Diseño de estrategias de m-learning. Desarrollo de aplicaciones*. Workshop en Investigación en Ciencias de la Computación, WICC. ISBN 978-950-605-570-7. Ushuaia. 2014.
28. Morales, M. I., Herrera, S. I., Fennema, M. C., Goñi, J. *Estrategias de m-learning para la enseñanza de la Matemática en carreras de Ingeniería*. VIII Congreso Argentino de Enseñanza de la Ingeniería. Tucumán. 2014.
29. Naismith, L., Lonsdale, P., Vavoula, G., and Sharples, M. *Literature Review in Mobile Technologies and Learning*. ISBN: 0-9548594-1-3. Univ. of Birmingham, 2004.
30. Pachler, N., Bachmair, B. & Cook, J. *Mobile Learning. Structures, Agency, Practices*. Ed. Springer, ISBN 978-1-4419-0584-0. New York, USA, 2010.
31. Pachler, N., Cook, J., Bachmair, B. *Appropriation of mobile Cultural resources for learning*. International Journal of Mobile and Blended Learning, 2(1), 1-21. 2010.
32. Quinn, C. N. *The Mobile Academy. mLearning for Higher Education*. Ed. Jossey-Bass. USA. 2012.
33. Rocabado Moreno, S., Herrera, S. I., Morales, M. I., Estellés, C. *M-learning en zonas de recursos limitados*. Congreso TE&ET. Santiago del Estero, 2013.
34. Sanz, C. y Otros. *Integración de la tecnología móvil a los entornos virtuales de enseñanza y de aprendizaje*. Publicado en TE&T.
35. Sanz, C., A. Zangara, C. Manresa-Yee. *E-activities in teaching processes using ICTs collaborative activity as a case study*. EDULEARN12 Proceedings, ISBN 978-84-695-3491-5, pp. 2034-2041. Barcelona, 2012.
36. Traxler, J. *Defining Mobile Learning*. IADIS International Conference M-Learning. 2005.
37. Traxler, J. *Defining, Discussing and Evaluating Mobile Learning: the moving finger writes and having write*. The International Review of Research in Open and Distance Learning. Vol.8, N°2. 2007.
38. Traxler, J. *Mobile Learning: Starting in the Right Place, Going in the Right Direction?*, International Journal of Mobile and Blended Learning, 3(2), 57-67, April-June 2011.
39. Vigotsky, L. *El desarrollo de los procesos psicológicos superiores*. Ed. Crítica. Barcelona. 1979.
40. Woodill, G. *The mobile learning edge*. Ed. Mc Graw Hill, 2011.

Desarrollo Dirigido por Modelos Basado en Componentes de Interfaz de Usuario

Doctorado en Ciencias Informáticas
Facultad de Informática - Universidad Nacional de La Plata

Autor: Pablo Martín Vera

GIDFIS (Grupo de Investigación, Desarrollo y Formación en Innovación de Software)
Departamento de Ingeniería e Investigaciones Tecnológicas Universidad Nacional de La
Matanza. pvera@unlam.edu.ar

Directora: Claudia Pons

LIFIA (Laboratorio de Investigación y Formación en Informática Avanzada), Facultad de
Informática, Universidad Nacional de La Plata. cpons@info.unlp.edu.ar

Co Directora: Carina González González

Universidad La Laguna, La Laguna, España. cjgonza@ull.es

Fecha de Exposición: 14 de Septiembre de 2015

Resumen

Esta tesis presenta una metodología de modelado para aplicaciones web móviles utilizando técnicas de desarrollo dirigido por modelos (MDD). Mediante la creación de sólo dos diagramas, un diagrama de datos y un diagrama de interfaz de usuario (que además incluye la navegación) es posible definir el comportamiento completo de una aplicación.

Por estar esta metodología basada en MDD incorpora dos transformaciones; la primera desde el modelo de datos a una versión inicial del modelo de interfaz de usuario, lo que reduce considerablemente el esfuerzo de modelado, ya que luego ese segundo modelo solo deberá ser adaptado a las necesidades particulares. La segunda transformación toma los modelos realizados y genera el código fuente completo, 100% funcional de una aplicación web móvil, además del script de la base de datos correspondiente. Ambos modelos están basados en una extensión conservativa de UML. El modelo de datos está basado en el diagrama de clases y el modelo de interfaz de usuario utiliza el diagrama de componentes de UML. Para poder especificar el comportamiento de la interfaz de usuario se definen una serie de componentes que a su vez pueden ser configurados con información tomada del modelo de datos. La configuración se basa en valores etiquetados propios para cada tipo de componente. Para facilitar el proceso de construcción de los modelos y su configuración se ha desarrollado una herramienta de soporte, que permite no solo modelar, sino también realizar las transformaciones establecidas en la metodología, obteniendo como resultado final una aplicación funcional sin escribir una sola línea de código.

Además esta tesis establece las ventajas de utilizar componentes configurables en el desarrollo dirigido por modelos, haciendo que el esfuerzo de programación se realice una única vez al establecer las transformaciones y que luego pueda ser aplicado a una amplia gama de aplicaciones de distintos dominios.

Palabras Clave: MDD, UML, COMPONENTES, MÓVIL, INTERFAZ DE USUARIO

1. Definición del Problema

El modelado de aplicaciones es un área subestimada por la industria de software. Especialmente en pequeñas y medianas empresas donde el modelado es considerado, en numerosas oportunidades, una pérdida de tiempo. En otros casos, los modelos solo se utilizan en etapas tempranas del desarrollo para la toma de requerimientos o como documentación inicial que luego no se actualiza con los cambios realizados durante el desarrollo de la solución, por lo que rápidamente quedan obsoletos. Por esa razón surge la idea de dar más importancia a los modelos. Los modelos pueden ser utilizados para generar de forma automática una aplicación o por lo menos parte de ella. “La automatización del proceso de desarrollo de software consiste en comenzar desde un alto nivel de representación de las características deseadas del software y derivar una aplicación ejecutable a partir de ella, posiblemente mediante un conjunto de pasos intermedios que permitan algún grado de interacción del usuario con el proceso de generación” [1]

El desarrollo de software mediante la construcción de modelos que permitan luego la generación automática de código fuente a partir de los mismos, es una tendencia iniciada hace varios años atrás. Puede encontrarse con diversos nombres: MDD (Model Driven Development), MDA (Model Driven Architecture), MDSE (Model Driven Software Engineering), MDE (Model Driven Engineering), entre otras. Sin embargo estas siglas tienen una categorización y relación entre ellas pero lo importante es que todas tienen algo en común, utilizan modelos y transformaciones.

Un modelo es una representación abstracta de la realidad, en este caso los modelos que utilizaremos son modelos software y por lo tanto podrán representar partes o componentes de un sistema. Un modelo puede tener una representación gráfica y también una descripción textual, lo importante es que siga ciertas reglas para su conformación.

Una transformación es un proceso que toma como entrada un modelo y genera como resultado otro modelo o código fuente. Por ejemplo el enfoque MDA de la OMG [2] [3], utiliza diferentes tipos de modelos con diferentes niveles de abstracción. Partiendo de modelos independientes de la plataforma (PIM) hasta llegar a modelos específicos para cada plataforma (PSM). “El PIM permite una representación visual del modelo, utilizando un nivel alto de abstracción. Los detalles de los modelos ambientales pueden ser expresados y precisados claramente en UML sin utilizar ningún formalismo particular de la implementación... un PSM se desarrolla mediante un mapeo de un PIM a una plataforma computacional y a un lenguaje de programación específico” [4].

Para llegar desde el PIM al PSM se deben realizar transformaciones automáticas o semi-automáticas. El objetivo final de estas técnicas es automatizar el proceso de generación del código fuente permitiendo que los diseñadores se centren en los modelos, más que en proceso de codificación.

Sin embargo, el principal problema, es que la mayoría de las técnicas existentes son difíciles de utilizar y requieren de un arduo proceso detallando modelos y configurando transformaciones para lograr obtener código fuente utilizable y aun así la mayoría de ellas solo permite generar parte del código fuente. Algunos autores comparten la visión de que las metodologías de modelado para generación de código fuente son demasiado complejas “Para realizar un programa útil partiendo de un modelo, el modelo debe ser tan complicado que la gente que no puede programar no lo puede comprender y los que pueden programar preferirán escribir el código antes que realizar los correspondientes modelos” [5].

Es por este motivo que surge la necesidad de crear una metodología de modelado fácil de utilizar, que no requiera de modelos complejos y que como resultado se pueda obtener el código fuente completo de una aplicación.

2. Hipótesis

Es posible desarrollar una metodología de modelado que incluya toda la información necesaria para generar el código fuente completo de una aplicación realizando transformaciones automáticas, con pocos modelos, simples y fácilmente entendibles.

3. Contribución Científica

- Utilización de componentes configurables para facilitar la generación de código fuente en MDD
- Diseño de una metodología para el modelado de aplicaciones web móviles centradas en los datos.

4. Solución Propuesta

Como solución se plantea el diseño de una metodología de modelado basada en componentes de interfaz de usuario configurables que incluye toda la información necesaria para generar el código fuente completo de una aplicación. Esta metodología se denomina CBMDD (Component Based Model Driven Development)

Los pasos de la metodología pueden verse en la Figura 1 y son los siguientes:

1. El diseñador realiza el modelo de datos del sistema.
2. Tomando como base el modelo de datos, la herramienta de transformación genera de forma automática una primera versión del modelo de interfaz.
3. El diseñador realiza los ajustes al modelo de interfaz, cambiando la configuración y/o agregando nuevos componentes.
4. Tomando los dos modelos terminados, se realiza la segunda y última transformación que genera el script de la base de datos y el código fuente de la aplicación.

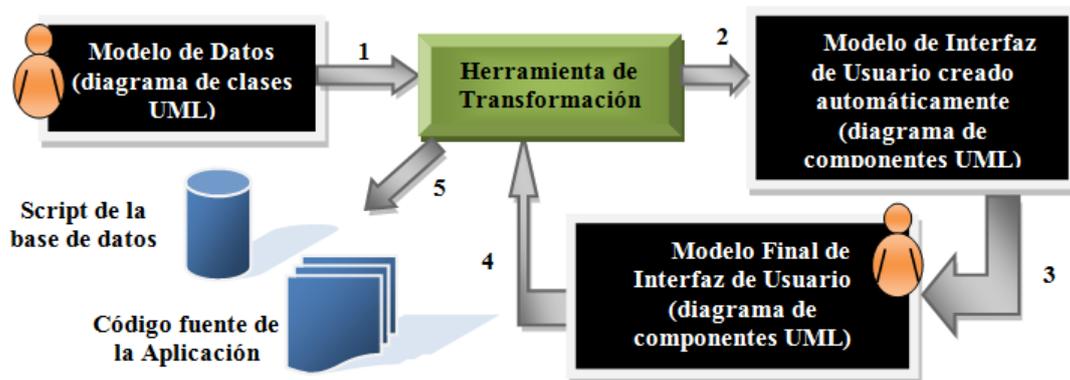


Fig 1. Pasos de la metodología de modelado

Ambos diagramas están basados en diagramas tradicionales de UML pero extendidos mediante estereotipos y valores etiquetados. Todas las extensiones realizadas se definen mediante un perfil de UML. Todos los nuevos estereotipos definidos heredan de alguna metaclassa de UML generando de esta forma una extensión conservativa. “Conservativa significa que los elementos de modelo del metamodelo de UML no se modifican... todos los nuevos elementos se relacionan por herencia hacia al menos un elemento del modelo del metamodelo de UML” [6].

4.1. Modelo de Datos

El primer paso es realizar el modelo de datos de la aplicación, para ello se utiliza el diagrama de clases de UML donde cada clase representa una tabla de la base de datos. En este diagrama se define información adicional que luego facilitará la creación de la base de datos y la visualización de información en el sistema. Esta información adicional se expresa en el diagrama de clases mediante estereotipos que permiten determinar la clave primaria de una clase utilizando el estereotipo Identifier y una descripción textual de dicha clase mediante el estereotipo Descriptor. Adicionalmente se utiliza el estereotipo enumeration ya presente en UML para definir por ejemplo distintos estados de un modelo.

4.2. Modelo de Interfaz de Usuario

En este modelo se diseñaran las distintas pantallas que conformaran el sistema, determinando además, cuál es el flujo de navegación entre dichas pantallas. Para modelar la interfaz de usuario se utilizará el diagrama de componentes de UML, donde cada componente se corresponderá con una pantalla del sistema.

Para identificar el tipo de pantalla que se desea modelar se crearon estereotipos que se aplican a los componentes de UML para tipificarlos. Los estereotipos disponibles son: Login, List, Search, Menu, CRUD y UpdateView.

Cada tipo de componente tiene definido una serie de valores etiquetados que permite configurar su comportamiento para adaptarlo a la aplicación que se desea modelar. Algunos de estos valores etiquetados son comunes para la mayoría de los componentes y otros son particulares para cada uno de ellos. El detalle de configuración de cada componente y de sus valores etiquetados puede verse en [7].

Existen algunas opciones de configuración que son comunes como ser un título general del componente, que se utilizará para mostrar al usuario en la pantalla que se encuentra y la barra de navegación que está presente en todos los componentes que se configura mediante una serie de links que llevan a otros componentes. Además cada componente incluye opciones de configuración propia que permiten definir cómo será la interfaz de usuario. A continuación se enumeran dichas opciones de configuración y su muestran pantallas generadas automáticamente por la herramienta de modelado de CBMDD.

- Login: Este componente permite la autenticación de un usuario en el sistema. Además de permitir configurar contra que tabla se realizará la autenticación, incluye la posibilidad de permitir que el usuario en lugar de ser ingresado, sea seleccionado, esto es útil en los dispositivos móviles ya que el ingreso de texto es una tarea dificultosa. La Figura 2 muestra dos pantallas de login, una en la cual el usuario debe ser ingresado y la otra donde directamente se selecciona.



Fig. 2. Pantallas generadas automáticamente a partir de componentes del tipo Login

- List: Este componente muestra listados de información. Su configuración permite: (a) Definir qué datos se van a mostrar; (b) Definir datos adicionales que se pueden mostrar en una segunda fila de información correspondiente a cada ítem. (c) Establecer filtros por defecto que se aplican al listado visualizado; (d) Determinar el orden del listado.

La Figura 3 muestra dos ejemplos de pantallas generadas a partir de componentes del tipo List. La imagen de la izquierda muestra un listado donde la información se presenta en varias líneas, una línea principal con un título y una línea adicional mostrando información complementaria, este caso con la categoría y fecha de vencimiento. En el lado derecho de la imagen se ve un listado con un solo dato mostrando la descripción de las categorías, pero en la barra de navegación además de la opción para volver al menú principal hay un botón para ir a la pantalla que permite agregar una nueva categoría.



Fig. 3. Pantallas generadas automáticamente a partir de componentes del tipo List

- Search: es un componente que permite mostrar un listado de información pero el mismo puede ser filtrado en tiempo de ejecución por el usuario. Incluye la misma configuración del componente List pero capacidad de definir sobre que campos se permitirá realizar la búsqueda y el tipo de control asociado al mismo. Los tipos de filtros disponibles a aplicar son: (a) Texto Libre; (b) Selección Simple; (c) Selección Múltiple; (d) Rango de Fechas; (e) Verdadero o Falso.

La Figura 4 muestra un ejemplo de los filtros de búsqueda que se generan en la interfaz de la aplicación.



Fig. 4. Pantalla de búsqueda generada automáticamente a partir de un componente Search

- **Menu:** este componente permite mostrar al usuario una pantalla con un menú. Su configuración es simple ya que solo se deben definir las opciones de dicho menú mediante una serie de links que apuntan a otros componentes tal como puede verse en la Figura 5.



Fig. 5. Pantalla generada automáticamente a partir de un componente del tipo Menu

- **CRUD:** es un componente para crear, modificar, visualizar y eliminar datos. Por defecto este componente solicita al usuario el ingreso de todas las propiedades de la clase que se está editando, pero incorpora una potente configuración que permite:
 - Definir valores por defecto al crear o actualizar un objeto
 - Evitar completar determinadas propiedades de una clase al crear o actualizar un objeto
 - Crear registros adicionales de otras tablas al crear o actualizar un objeto, lo que es muy útil para crear tablas de log.
 - Definir propiedades opcionales

La Figura 6 muestra una pantalla de edición donde para cada propiedad se generan automáticamente los controles adecuados para que el usuario pueda ingresar los datos. Puede verse que para clases relacionadas se genera un control de selección con la descripción de la otra entidad como es el caso de la categoría.

Fig. 6. Pantalla de edición generada automáticamente a partir de un componente del tipo CRUD

- UpdateView: este componente permite realizar operaciones de actualización sobre objetos incorporando características especiales que lo diferencian del CRUD. En muchos sistemas es posible que un único objeto deba ser actualizado parcialmente donde solo algunas de sus propiedades se deban completar y otras sean mostradas en modo de solo lectura. Este control es muy útil para sistemas donde un objeto va pasando por diferentes estados y en cada estado se registran distintos valores. Para actualizar cada uno de esos estados podría configurarse un componente UpdateView diferente. La Figura 7 muestra un ejemplo de una pantalla generada por un componente UpdateView donde se configuraron 3 propiedades de solo lectura (título, categoría y fecha de vencimiento) y solo se permite editar al usuario la descripción. En este caso además al actualizar los datos el estado de la tarea cambia automáticamente a Completa.

Fig. 7. Pantalla generada automáticamente a partir de un componente del tipo UpdateView

4.3. Herramienta de Modelado

La metodología de modelado fue desarrollada basada en UML utilizando una extensión conservativa mediante estereotipos y valores etiquetados, esto hace que sea posible crear los modelos en cualquier herramienta basada en UML ya que los mecanismos de extensión están presentes en todas ellas. Sin embargo la configuración de cada valor etiquetado es particular y requiere recordar la sintaxis para realizar una configuración correcta. Este es un primer motivo por el cual se decidió desarrollar una herramienta que permita facilitar dicha configuración, el segundo motivo es que dicha herramienta dará soporte además a las transformaciones entre modelos.

La herramienta es una aplicación Web desarrollada en C# bajo el framework Microsoft ASP.NET V4 utilizando Web Forms [8]. La misma fue desarrollada con los siguientes objetivos:

1. Facilitar el modelado evitando que el diseñador deba recordar la configuración particular de cada componente.
2. Permitir la interoperabilidad con otras aplicaciones de modelado mediante XMI.
3. Validar la correcta configuración de los componentes en el caso de que se realice con una herramienta externa.
4. Realizar las transformaciones automáticas previstas en la metodología.

Además la herramienta fue desarrollada para ser escalable permitiendo:

- Agregar fácilmente nuevos componentes.
- Permitir reutilizar controles para configurar los componentes.
- Poner a disposición clases para generar fácilmente código fuente en cualquier lenguaje.

5. Resultados

Se desprende de las secciones anteriores que mediante la herramienta construida es posible realizar los modelos y aplicar las transformaciones. Generándose los scripts de la base de datos y el código fuente de la aplicación. Por otra parte al comenzar un proyecto la herramienta muestra los estados de los diagramas (por ejemplo: diagrama de clases – sin datos; Generar Componentes – no generado; etc.). Estos estados cambian a medida que se trabaja con la herramienta. También permite evitar por ejemplo que el usuario cambie el diagrama de clases y no vuelva a generar el de componentes. Permitiendo que exista una trazabilidad entre los modelos y transformaciones aplicadas (ver figura 8).

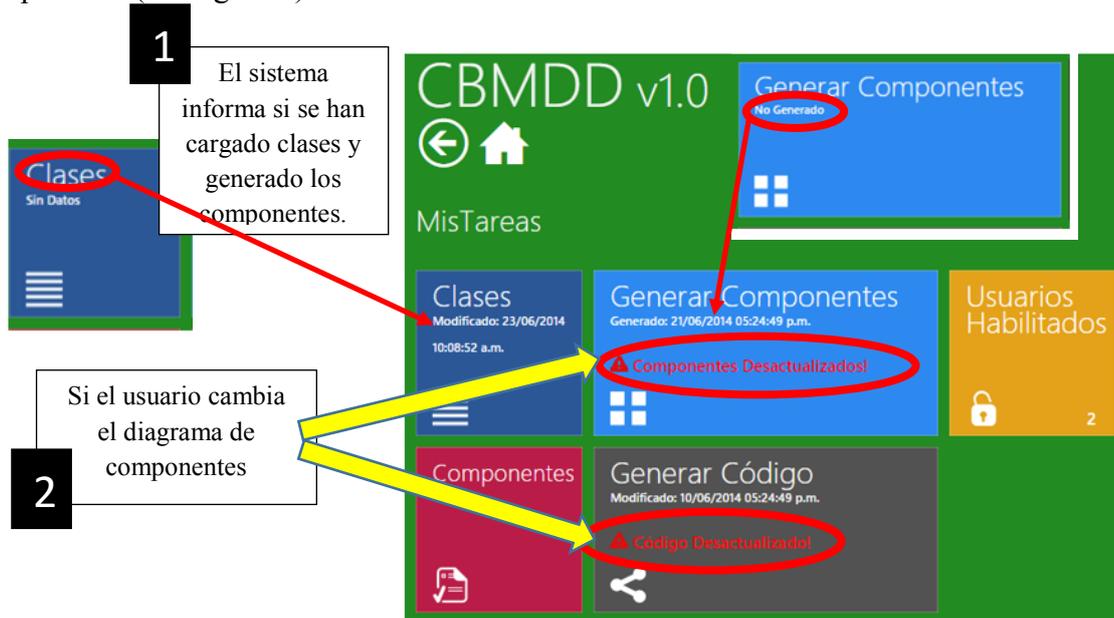


Fig.8. Menú de acciones de un proyecto de modelado

Se han realizado diversas aplicaciones con distintos grados de complejidad pudiendo generarse el código fuente funcional de ellas a través de la herramienta, también se ha trabajado en la interoperabilidad permitiendo que pueda exportarse el XMI del modelo y luego importarse en otra herramienta de modelado. Para esto se creó un parser de XMI 2.2 que brinda interoperabilidad con el EA (Enterprise Architecture) permitiendo importar en la herramienta los modelos previamente construidos EA o exportarlos al EA.

Las opciones para importar y exportar modelos mediante XMI se agregaron tanto para el modelo de datos como para el modelo de componentes. En el caso del proceso de importación también se realiza una validación para comprobar que el modelo que se está importando este correctamente configurado.

6. Conclusiones

El empleo de MDD permite poner el foco de atención en las primeras etapas haciendo que los modelos sean la parte más importante, pudiendo transformarse y evolucionar hasta obtener el código fuente de la aplicación. Es claro que el esfuerzo entonces no está puesto en la programación sino en el diseño. En contrapartida, algunas metodologías de MDD requieren realizar gran cantidad de modelos con inter-relaciones complejas, en donde el esfuerzo final utilizado en las etapas de modelado termina siendo equiparado con el que se requeriría para desarrollar una aplicación de cero de forma tradicional. La presente tesis, plantea la importancia de centrarse en los modelos y como principal contribución teórica se construye una metodología que extiende de forma conservativa a

UML, mediante la cual con 2 modelos es posible efectuar transformaciones automáticas de forma transparente al usuario final y obtener el código fuente funcional de una aplicación. Esta metodología se planteó considerando diversos aspectos: (1) reutilización de componentes, (2) facilidad en el modelado, (3) extensión conservativa de uml y (4) completitud de la información.

Además se desarrolló una herramienta que permite implementar la metodología efectivamente, chequear la trazabilidad entre los modelos generados, importar modelos desde otra herramienta de modelado facilitando la interoperabilidad, efectuar las transformaciones automáticas y generar finalmente como resultado tanto los scripts de la base de datos como el código fuente de la aplicación.

La herramienta desarrollada facilita aún más la tarea de modelado ya que evita que el diseñador deba lidiar con la sintaxis concreta de la configuración de los componentes brindando interfaces gráficas de configuración y generación de los modelos. Por otra parte, incorpora los procesos de transformación necesarios en la metodología.

Cabe destacar que la herramienta de soporte fue desarrollada en forma modular permitiendo rápidamente la incorporación de nuevos componentes y nuevos valores etiquetados para permitir la expansión futura de la metodología. También brinda clases para facilitar la tarea de la programación de las transformaciones para generar código fuente incorporando el concepto de templates lo que permite generar con un mismo modelo, código para distintas plataformas.

Si bien la metodología CBMDD ha sido planteada para dispositivos móviles es posible adaptarla a otras plataformas. En el caso de que la interfaz de usuario que se desee modelar tenga una distribución de pantalla más compleja es posible agregar nuevos componentes así como también agregar parámetros de configuración a los componentes existentes según las necesidades del caso.

7. Trabajos Futuros

- Creación de nuevos templates para generar interfaces para distintas plataformas y distintos lenguajes.
- Incorporar opciones de interoperabilidad con otras herramientas de modelado.
- Hacer más poderosos los templates de creación de código fuente incorporando parámetros a los mismos para generar interfaces con distintas opciones sin necesidad de crear un nuevo template. Por ejemplo podría seleccionar si la navegación primaria es persistente (siempre fija) o transitoria (se despliega un menú con una opción) en la pantalla. O incluso permitir seleccionar imágenes para que el menú principal sea icónico en lugar de textual (patrón Springboard).
- Incorporación de nuevos controles o nuevas configuraciones para hacer más poderoso el modelo.

8. Publicaciones Efectuadas

1. Revista Colombiana de Computación (RCC) “La interfaz de usuario como punto de partida para la creación automática de aplicaciones móviles – Un enfoque basado en MDD”. 2015
2. International Journal of Information Technologies and Systems Approach (IJITSA) - Volúmen 8, Número 2, pp 80 – 100. “Component Based Model Driven Development – An Approach for Creating Mobile Web Applications from Design Models”. Special Issue on HCI 2015
3. XX Congreso Argentino de Ciencias de la Computación (CACIC). “Automatic Creation of Mobile Web Applications from Design Models“. Universidad Nacional de La Matanza, San Justo, Buenos Aires, Argentina. 2014
4. XVI Workshop de Investigadores en Ciencias de la Computación (WICC). “Generación Automática de Aplicaciones Web Móviles Mediante Componentes Configurables“. Universidad Nacional de Tierra del Fuego, Ushuaia, Tierra del Fuego, Argentina. 2014

5. Third International Conference on Software and Emerging Technologies for Education, Culture, Entertainment, and Commerce (SETECEC): New Directions in Multimedia Mobile Computing, Social Networks, Human-Computer Interaction and Communicability. “Tool for developing Mobile Web Application from UI Models – Based on CBHDM Methodology”. Venecia, Italia 2014
6. XIX Congreso Argentino de Ciencias de la Computación (CACIC). “Modeling Complex Mobile Web Applications from UI Components – Adding Different Roles and complex Database Design”. Universidad CAECE, Mar del Plata, Buenos Aires, Argentina 2013.
7. XLII Jornadas Argentinas de Informática (42 JAIIO). “Metodología de Modelado de Aplicaciones Web Móviles Basada en Componentes de Interfaz de Usuario”. Facultad de Matemática, Astronomía y Física de la Universidad Nacional de Córdoba, Córdoba Capital, Córdoba, Argentina. 2013
8. XV Workshop de Investigadores en Ciencias de la Computación (WICC). “Metodología de Modelado de Aplicaciones Web Móviles Basada en Componentes”. Universidad Autónoma de Entre Ríos (UADER), Paraná, Entre Ríos, Argentina. 2013
9. XVII Congreso Argentino de Ciencias de la Computación (CACIC). “MDA based Hypermedia Modeling Methodology using reusable components”. Universidad Nacional del Sur, Bahía Blanca, Buenos Aires, Argentina. 2012
10. XV Workshop de Investigadores en Ciencias de la Computación (WICC) - “Utilizando el Enfoque MDA para la Construcción de Aplicaciones Web Móviles Centradas en los Datos”. Universidad Nacional de Misiones, Posadas, Misiones, Argentina. 2012
11. VII Congreso Colombiano de Computación (7CCC). “User Interface and Navigation Modeling Methodology for Mobile Hypermedia Systems”. Medellín, Colombia. 2012

Referencias

- [1] Brambilla, M. C. (2012). *Model-driven software engineering in practice* (Vol. 1). Synthesis Lectures on Software Engineering.
- [2] Kleppe, A., Warmer, J., & Wim, B. (2003). *MDA explained: the model driven architecture: practice and promise*. Addison-Wesley Professional.
- [3] OMG. (2003). *MDA Guide Version 1.0.1*. Retrieved 11 04, 2014, from <http://www.omg.org/cgi-bin/doc?omg/03-06-01.pdf>
- [4] Papajorgji, P., Beck, H. W., & Braga, J. L. (2004). n architecture for developing service-oriented and component-based environmental models. *Ecological Modelling*, 179(1), 61-76.
- [5] Steimann, F., & Kühne, T. (2005, 12). Coding for the Code. (ACM, Ed.) *Queue - Managing Megaservice*, 3(10), 44-51.
- [6] Rossi, G., et al. *Web Engineering: Modelling and Implementing Web Applications*. 2008.
- [7] Vera P., Pons C. Gonzales C, Giulianelli D., Rodriguez R. *MDA based Hypermedia Modeling Methodology using reusable components*. s.l. : XVIII Congreso Argentino de Ciencias de la Computación, 2012.
- [8] Microsoft. *ASP.NET Web Forms*. [Online] [Cited: 01 05, 2015.] <http://www.asp.net/web-forms>.

Representación y razonamiento sobre las decisiones de diseño de arquitectura de software.

Autor: María Celeste Carignano

Director: Dr. Horacio Leone

Codirector: Dr. Silvio Gonnet

Fecha de exposición: 1 de Julio de 2015

Tesis presentada para la obtención del grado de Doctor en Ingeniería, mención Sistemas de Información, Facultad Regional Santa Fe, Universidad Tecnológica Nacional.

1. Introducción

La arquitectura de un sistema puede ser vista como las estructuras del sistema que comprenden a los elementos de software, las propiedades visibles externamente de dichos elementos, y las relaciones entre ellos [1]. En las últimas décadas las arquitecturas de software han ido tomando un protagonismo creciente dentro del ciclo de vida de un proyecto de software. Este protagonismo se fue incrementando de la mano de la complejidad de los sistemas construidos.

En la actualidad, construir sistemas sin diseñar previamente su arquitectura es equivalente a construir edificios sin definir previamente sus planos.

La importancia de las arquitecturas de software no solo radica en que permiten capturar la o las estructuras de un sistema, sino que también ([1], [2]):

- son un vehículo de comunicación entre los "stakeholders",
- sirven como un mecanismo de educación y entrenamiento,
- manifiestan las decisiones tempranas sobre los sistemas,
- se constituyen como abstracciones transferibles de sistemas diseñados,
- sirven como base para el análisis de los sistemas y sus construcciones,
- facilitan el razonamiento frente a cambios en los sistemas, y
- permiten predecir características de calidad del sistema a partir de su estudio.

Si bien el diseño de arquitecturas de software tiene un costo asociado al tiempo, el esfuerzo y los recursos utilizados para tal fin, el no contar con una arquitectura completamente diseñada tiene un costo mayor [2] e impacta en la calidad del producto resultante, el tiempo insumido para el desarrollo del sistema e incluso en la culminación con éxito del proyecto de desarrollo.

El diseño de arquitecturas de software es un proceso altamente creativo que aún no ha sido formalizado, por lo que las actividades llevadas a cabo para construir la arquitectura de un sistema son aquellas que los arquitectos involucrados consideran convenientes y pertinentes según el método de diseño que utilicen, su experiencia, conocimientos y habilidades personales. Según Kruchten [3] toda arquitectura de software se crea a partir de tres fuentes:

- método: éste puede ser visto como una manera sistemática, concisa y documentada mediante la cual la arquitectura es derivada desde los requerimientos del sistema y las restricciones tecnológicas;
- intuición: considerada como la habilidad de concebir sin razonamiento consciente;
- reutilización: la mayoría de los elementos de una arquitectura de software son obtenidos de otras arquitecturas, principalmente cuando los arquitectos se encuentran familiarizados con un sistema previo del mismo tipo, otro sistema con características similares, o alguna arquitectura encontrada en la literatura técnica.

Kruchten [3] opina que la proporción en que estas tres fuentes influyen en un diseño arquitectónico varía de acuerdo a la experiencia del arquitecto y al grado de novedad del sistema a ser diseñado. Por ejemplo, en sistemas clásicos la reutilización de soluciones prima por sobre la intuición, ya que se tienen como referencias las arquitecturas definidas para sistemas con los cuales se ha trabajado previamente (Fig. 1 (a)). Mientras que en sistemas sin precedentes, la intuición juega un papel importante (Fig. 1 (b)). Sin embargo, en ambos casos el proceso de construcción de

la arquitectura debe estar soportado, en mayor o menor medida, por algún método de diseño validado. Las arquitecturas que se construyen sin emplear un método están destinadas al fracaso.

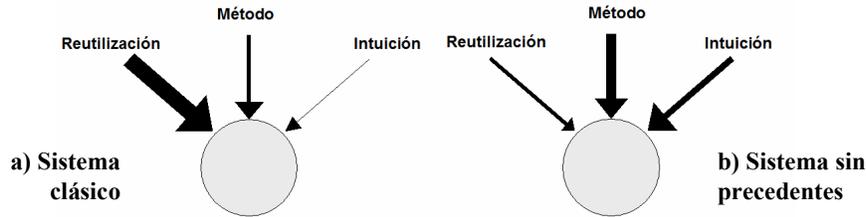


Figura 1: Fuentes de un diseño arquitectónico, extraído de [3]

De alguna manera, la esencia del trabajo de un arquitecto de software es encontrar y aplicar un balance correcto entre estas tres fuentes de una arquitectura [4].

En lo que respecta a métodos de diseño, el arquitecto cuenta con una amplia variedad de ellos. En las últimas décadas se han definido y documentado diferentes métodos, procesos y técnicas de diseño de arquitecturas de software. Debido a que muchos de ellos fueron desarrollados de manera independiente, difieren en diversos aspectos, incluso en los términos que utilizan.

Hofmeister y otros [5] efectuaron una comparación de cinco de los métodos de diseño de arquitecturas utilizados en la industria de software. A partir de esta comparación, se identificaron varios puntos en común entre ellos y algunas diferencias. Como principal aporte resultante de dicho trabajo se definió un proceso de diseño arquitectónico genérico que describe las actividades llevadas a cabo por los arquitectos: análisis, síntesis, y evaluación arquitectónica (Fig. 2). Estas actividades no se realizan de manera secuencial sino que forman parte de un proceso iterativo en el cual los artefactos van evolucionando a medida que éstas son ejecutadas.

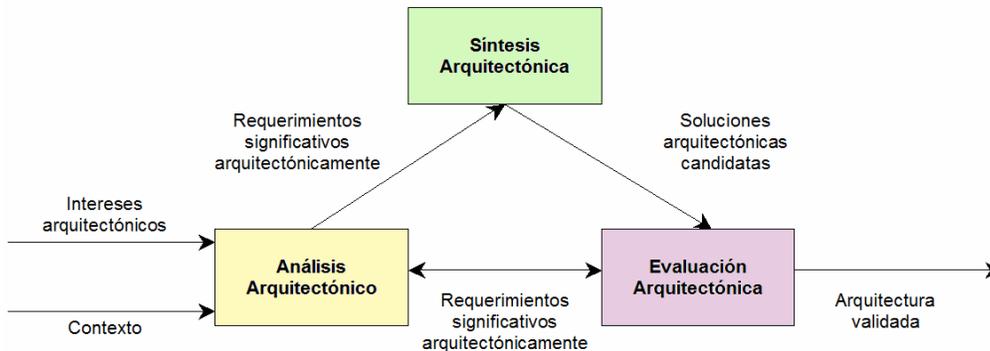


Figura. 2: Actividades del diseño arquitectónico, extraído de [5].

Debido a que no es posible analizar, encontrar soluciones y evaluar la arquitectura para todos los intereses de todos los “stakeholders” de manera simultánea, es que se divide el problema y se comienza trabajando sobre una parte del mismo, según los arquitectos lo consideren conveniente. Se lleva a cabo una secuencia iterativa e incremental de actividades de análisis-síntesis-evaluación. En cada iteración el arquitecto puede proponer y aplicar una o mas soluciones, las cuales pueden ser soluciones nuevas e innovadoras o bien soluciones ya conocidas que son adaptadas al nuevo problema para su reutilización. Ambas son el resultado de actividades altamente cognitivas por parte del arquitecto.

La reutilización o reuso es el proceso en el cual productos de trabajo de software existentes (como ser código, documentación, diseño, datos de prueba, herramientas y especificaciones) son utilizados en nuevos desarrollos, preferentemente con modificaciones mínimas [6]. En este contexto, el arquitecto debe encontrar alguna analogía entre parte o la totalidad del problema planteado para el sistema que intenta construir y los problemas que originaron sistemas ya existentes que él conoce. La analogía es un proceso cognitivo sofisticado en el cual dos situaciones,

una origen y una destino, son analizadas para encontrar patrones estructurales comunes ([7], [8]) con el objetivo de aplicar el conocimiento disponible acerca de la situación origen en la situación destino una vez que se hayan efectuado las adaptaciones necesarias.

En la práctica, en una encuesta realizada en el marco de esta Tesis [9], se pudo observar que el 98.5% de los encuestados hacía uso de la reutilización durante el diseño arquitectónico. Raras veces los arquitectos comienzan a trabajar a partir de una hoja en blanco ([10], [11]).

2. Identificación del problema

En la actualidad, dentro del contexto del diseño arquitectónico existen varias herramientas que brindan asistencia a los arquitectos para la realización de sus tareas. Algunas permiten modelar los elementos de las descripciones arquitectónicas (entre las que se incluyen los ADLs siendo UML el más frecuentemente utilizado) y otras además brindan la posibilidad de documentar y gestionar el conocimiento arquitectónico subyacente a los diseños creados.

Estas herramientas permiten que el arquitecto capture elementos pertenecientes a la arquitectura diseñada y pueda recuperarlos y visualizarlos posteriormente para su evolución o evaluación. Sin embargo, a pesar del importante papel que juega la reutilización en el diseño de una arquitectura, estas herramientas no tienen un mecanismo o procedimiento definido que le brinde asistencia al arquitecto en la realización y sistematización de dicha práctica. Es decir, el arquitecto no cuenta con un soporte que le facilite la realización del razonamiento analógico que se lleva a cabo durante la reutilización de un diseño arquitectónico. No ocurre lo mismo en etapas de diseño más avanzadas en donde se detalla la forma en la que la arquitectura de un sistema de software será implementada. Ejemplos de esto se describen en [12] y [13].

3. Objetivos de la Tesis

El objetivo general de la Tesis está directamente vinculado a brindar soporte al arquitecto de software durante el diseño arquitectónico. Partiendo de la hipótesis de que los arquitectos de software reutilizan soluciones conocidas en nuevos diseños, se identifica la necesidad de contar con una herramienta, tanto conceptual como computacional, que pueda colaborar con los arquitectos de software en dos aspectos:

- uno individual, que se encuentra relacionado con facilitar a los arquitectos la recuperación de experiencias pasadas propias, haciéndolas accesibles de manera de que puedan mapearlas con los problemas actuales para dar solución al diseño de un sistema;
- uno grupal, vinculado con facilitar a los arquitectos la recuperación de experiencias pasadas de otros arquitectos, haciéndolas accesibles de manera de que puedan mapearlas con los problemas actuales para dar solución al diseño de un sistema.

4. La propuesta

Para alcanzar el objetivo mencionado, se propone la aplicación de la técnica de Razonamiento Basado en Casos en el ámbito del diseño arquitectónico. Razonamiento Basado en Casos ha sido caracterizado como una instancia particular del razonamiento analógico en el cual se dan analogías intra-dominios, en lugar de inter-dominios [14]. Difiere de otros enfoques de Inteligencia Artificial debido a que enfatiza el rol de la memoria en el proceso cognitivo del ser humano. Su aplicación favorece el aprendizaje por experiencia, dado que generalmente es más sencillo aprender reteniendo la experiencia de resolver un problema concreto que obtener generalizaciones a partir de un problema [15].

La utilidad esperada de la propuesta presentada involucra: (i) la mejora en la calidad del producto de software resultante, y (ii) la reducción de los costos de los proyectos de software debido a la disminución del esfuerzo empleado por los arquitectos para diseñar las arquitecturas correspondientes.

Para un único individuo, es difícil manejar mentalmente la complejidad de la totalidad de los sistemas actuales cuando estos son grandes y deben satisfacer muchos requerimientos y restricciones. Al contar con una herramienta que le facilite el recuerdo de experiencias pasadas similares, el arquitecto podrá:

- acceder a información de soluciones de diseños arquitectónicos llevados a cabo en el pasado y analizar las decisiones tomadas para determinar si pueden ser replicadas en el sistema actual. Esta información puede provenir de experiencias propias o de otros arquitectos nutriendo el conjunto de posibles soluciones a emplear,
- evitar la “vaporización del conocimiento”, y recordar sistemas o soluciones que por algún motivo pueden ser olvidadas o confundidas,
- contar con información proveniente de la experiencia acerca de cuales fueron las mejores soluciones aplicadas en el pasado, y cuales no deberían ser replicadas por haber considerado que su aplicación no fue adecuada.

Estos tres puntos describen el impacto en el esfuerzo llevado a cabo por los arquitectos durante el diseño arquitectónico, potenciando el tercero la disminución del esfuerzo por “retrabajo” al colaborar para que no se apliquen soluciones clasificadas como erróneas en situaciones anteriores que son consideradas similares a las nuevas. Además, el poder contar con información relacionada a la calidad de los sistemas diseñados con anterioridad propicia a mejorar la calidad resultante de los productos de software diseñados.

Como se pudo concluir a partir del análisis del trabajo de Tang y otros [16] y de la encuesta realizada en el marco de esta Tesis, durante un diseño arquitectónico, uno de los recursos más limitados para los arquitectos es el tiempo, por lo tanto para que la herramienta sea de utilidad:

- deberá ser fácil de usar para no entorpecer el trabajo de los arquitectos, y
- no deberá requerir mayor esfuerzo del que lleva documentar correcta y completamente una arquitectura, considerando que dentro de la definición de arquitectura también se incluye el conocimiento arquitectónico involucrado.

5. Descripción de la solución

Razonamiento Basado en Casos es un paradigma de resolución de problemas [15] que involucra el uso de experiencias pasadas para comprender y resolver nuevas situaciones [17]. Un caso denota una situación experimentada previamente, la cual ha sido capturada y aprendida de manera que pueda ser reutilizada en la resolución de problemas futuros [15]. Suele estar compuesto por [18]:

- el problema: que describe el estado del mundo cuando ocurre el caso,
- la solución: que establece la solución encontrada, y/o
- la salida: que describe el estado del mundo luego de que ocurrió el caso.

Esta técnica involucra la ejecución de cuatro actividades conocidas como las *4Res* (Recuperar, Reusar, Revisar y Retener), las cuales son presentadas en la Fig. 3.

La primera actividad es la de **RECUPERACIÓN** de los casos más similares. Comienza con la descripción de un problema (conocido como nuevo caso) y finaliza con la obtención de uno o más casos recuperados. Cada caso recuperado propone soluciones previamente aplicadas para resolver un problema similar. Involucra la evaluación de todos los casos en memoria y la determinación del grado de similitud entre cada uno de ellos y el nuevo caso. Generalmente es llevada a cabo en dos pasos [17]:

- primero se recupera un conjunto de casos candidatos. Los casos candidatos son casos con potencial para que a partir de ellos se efectúen predicciones relevantes sobre el nuevo caso,
- luego, se realiza un procesamiento para seleccionar el o los mejores casos del conjunto de casos candidatos previamente recuperados.

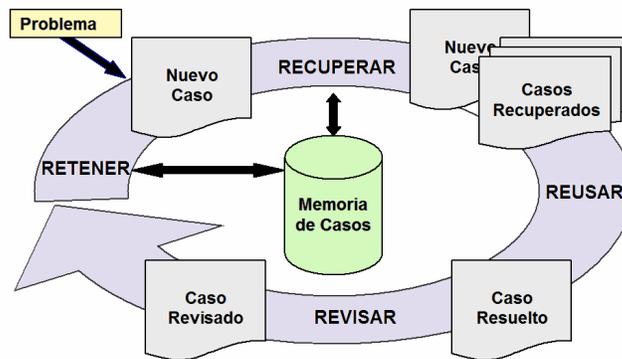


Figura 3: Actividades del proceso de Razonamiento Basado en Casos. Figura adaptada de [15].

La segunda actividad tiene como objetivo **REUSAR** la información y el conocimiento del caso recuperado para resolver el nuevo problema. Es la responsable de proporcionar una solución (conformando un caso resuelto) para un nuevo problema a partir de las soluciones del caso recuperado: soluciones antiguas son utilizadas como inspiración para resolver nuevos problemas [17].

La tercera actividad es la de **REVISIÓN** de la solución propuesta. Una vez que el nuevo caso ha sido resuelto debe ser probado en el mundo real. Cuando la solución generada por la actividad de reuso no es correcta surge una oportunidad para aprender de las fallas. De esta forma, durante la actividad de revisión, se llevan a cabo dos tareas [15]: la evaluación del caso resuelto generado por la actividad de reuso, y la reparación del caso resuelto, si es necesario, utilizando conocimiento específico del dominio.

Durante última actividad, la de **RETENCIÓN**, el caso revisado es incorporado a la memoria de casos como un caso aprendido con el objetivo de que esté disponible si un nuevo problema arriba para comenzar con el ciclo nuevamente.

Es importante aclarar que las actividades de reuso, revisión y retención raras veces ocurren sin la intervención humana [18].

5.1. Representación de casos en el contexto del diseño arquitectónico

Un razonador basado en casos es fuertemente dependiente de la estructura y el contenido de sus colecciones de casos [15]. En el contexto de esta Tesis, un caso describe la arquitectura diseñada para un sistema de software en particular y es llamado *caso arquitectónico*. A diferencia de la mayoría de las aplicaciones de Razonamiento Basado en Casos, en esta Tesis se utiliza un enfoque estructural [19] en donde los casos no se encuentran organizados como pares objeto-valor, sino que han sido definidos con un modelo de objetos complejo. Este modelo es representado utilizando diagramas de clases de UML ([20]) y es enriquecido mediante el empleo de OCL ([21]) para expresar restricciones definidas sobre el mismo.

Como puede observarse en la Fig. 2, cuando se debe diseñar la arquitectura de un sistema de software, la mayoría de la información con la que se cuenta está dada por los intereses arquitectónicos de los “stakeholders” y el contexto, o entorno, en el cual se inserta el sistema (ya sea operacional, político, organizacional, de negocio, etc.). Ambos conceptos ayudan a describir el problema asociado a un caso arquitectónico en particular. Por otro lado, una vez que el caso arquitectónico ocurre, es decir, luego de que la arquitectura del sistema ha sido diseñada, se obtiene como salida una arquitectura validada compuesta de elementos arquitectónicos junto con el razonamiento asociado a las decisiones de diseño tomadas por los arquitectos.

En esta Tesis, los casos arquitectónicos han sido estructurados en base al estándar internacional ISO/IEC/IEEE 42010:2011 [22] que define un marco conceptual estableciendo los términos y conceptos relacionados con el contenido y el uso de las descripciones arquitectónicas.

Dada la generalidad de los conceptos involucrados en el estándar ISO/IEC/IEEE 42010:2011, algunos de ellos han sido especializados para poder cumplir con el objetivo planteado previamente.

De esta forma, un caso arquitectónico se compone de un problema y una solución.

El *problema* identifica las necesidades que el sistema debe satisfacer mediante el empleo de restricciones de diseño y una representación particular de los requerimientos de calidad.

En una etapa temprana de la definición de una arquitectura de software, los requerimientos más relevantes son aquellos asociados con la calidad del sistema de software a diseñar, los cuales son conocidos como requerimientos de calidad. Un requerimiento de calidad describe alguna característica o atributo de calidad que la solución de software debe poseer, como ser tiempo de respuesta rápido, facilidad de uso o alta confiabilidad [23].

Debido a los problemas de ambigüedades, inexactitudes e inconsistencias [24] que presenta el lenguaje natural, éste no puede ser empleado como herramienta para especificar los requerimientos en el contexto de esta Tesis. En cambio, se propone la utilización de *escenarios de atributos de calidad* (adoptados del trabajo de Bass y otros [1]) para documentar de manera organizada los requerimientos de calidad de los "stakeholders". Dos tipos de escenarios son empleados: los escenarios generales y los escenarios específicos del sistema.

Los escenarios generales describen cómo una arquitectura permitiría responder a ciertos estímulos representando atributos de calidad [25]. Son genéricos e independientes del sistema y proveen un "framework" para formar un gran número de escenarios específicos del sistema [1]. Debido a su generalidad, pueden ser utilizados para crear los escenarios específicos de todos los sistemas diseñados. Los escenarios específicos del sistema expresan los requerimientos de calidad.

La representación empleada de escenarios fue complementada con la especificación de restricciones de diseño. Las *restricciones de diseño* son limitaciones sobre el diseño arquitectónico establecidas por los "stakeholders" del sistema. Estas restricciones acotan la libertad de los arquitectos al momento de tomar decisiones, imponiendo o impidiendo la utilización de determinadas estrategias de diseño, como ser tácticas o estilos arquitectónicos.

Los arquitectos utilizan estilos y tácticas arquitectónicas como ayuda en el proceso de diseño. Ambos están destinados a mejorar las decisiones de software de arquitectura y simplificar el proceso de diseño [26]. Las tácticas arquitectónicas están focalizadas en los atributos de calidad, son decisiones de diseño que ayudan a lograr una respuesta específica de un atributo de calidad en particular [27] por medio de la manipulación de algunos aspectos del modelo de atributo de calidad. Los estilos arquitectónicos son soluciones bien conocidas a problemas comunes de diseño [1]. Un estilo arquitectónico es una especialización de tipos de elementos y relaciones junto con un conjunto de restricciones sobre cómo pueden ser utilizados [2].

En principio, las restricciones de diseño limitan el espacio de soluciones disponibles. Frente a un conjunto de restricciones establecidas, es responsabilidad del arquitecto de software analizar y decidir si puede construir un sistema que las satisfaga. Para ello, es útil que cuente con información adicional como ser el grado de necesidad del cumplimiento de las restricciones (si son obligatorias u opcionales). Esta información cobra relevancia cuando las estrategias a aplicar para cumplir dos o más restricciones de diseño presentan conflictos entre si.

En la *solución* de un caso arquitectónico se describe la resolución del problema en términos que faciliten su reutilización en nuevos casos. Para determinar los elementos del diseño arquitectónico a reutilizar, es necesario tener en cuenta que existen dos niveles de reuso posibles [28]: (i) el reuso de las ideas y el conocimiento, y (ii) el reuso de los artefactos de diseño particulares y componentes. Debido a que la solución propuesta a los arquitectos debe ser de utilidad para el diseño de la nueva arquitectura, en el contexto de esta Tesis, la reutilización se lleva a cabo en el primer nivel: el del conocimiento. Los motivos que justifican esta elección están relacionados con dos hechos:

- Por un lado, algunas veces la correcta identificación de los artefactos que conforman el diseño arquitectónico depende del dominio específico de la aplicación. Teniendo en

cuenta esto, brindar a los arquitectos una solución para el nuevo caso basada en los artefactos y componentes de un caso recuperado de la memoria implicaría encontrar correspondencias entre los dominios de ambos casos. Debido a que uno de los requerimientos asociados a la propuesta presentada en esta Tesis, establece que no se debería requerir mayor esfuerzo por parte del arquitecto del que lleva documentar correcta y completamente una arquitectura, obligar a los arquitectos a documentar las arquitecturas con el suficiente nivel de detalle que permita establecer correspondencias entre dos dominios de aplicaciones aleatorios, violaría dicho requerimiento. Por lo tanto, para definir la estructura de la solución de un caso arquitectónico, se considera suficiente que los arquitectos puedan conocer cuales fueron las soluciones planteadas en los casos recuperados de manera genérica.

- Por el otro lado, contar únicamente con la identificación de los artefactos y componentes puede resultar inútil si no se acompaña de información vinculada a su uso. Por ejemplo, que el arquitecto del nuevo sistema conozca que para satisfacer un requerimiento de modificabilidad se utilizó un componente que representa al mapeador objeto-relacional “Hibernate” no es un aporte significativo. En cambio, es útil que el arquitecto del nuevo sistema conozca que para satisfacer un requerimiento de modificabilidad se decidió aplicar una táctica de separación de responsabilidades, motivo por el cual se utilizó un componente que representa al mapeador objeto-relacional “Hibernate”.

Desde esta perspectiva, los conceptos seleccionados para llevar a cabo la reutilización del conocimiento de los arquitectos son los de decisiones de diseño arquitectónico y estrategias de diseño. Ambos conceptos forman parte de la solución de un caso arquitectónico. Las decisiones de diseño arquitectónico son los elementos que establecen cuales fueron los objetivos del arquitecto durante la construcción de las descripciones arquitectónicas y de que manera cumplieron dichos objetivos.

5.2. Recuperación de las decisiones de diseño

La aplicación de la técnica de Razonamiento Basado en Casos para un caso en particular comienza con la recuperación desde la memoria de los casos más similares. En la presente Tesis, esta actividad involucra la realización de un proceso complejo, esquematizado en la Fig. 4.

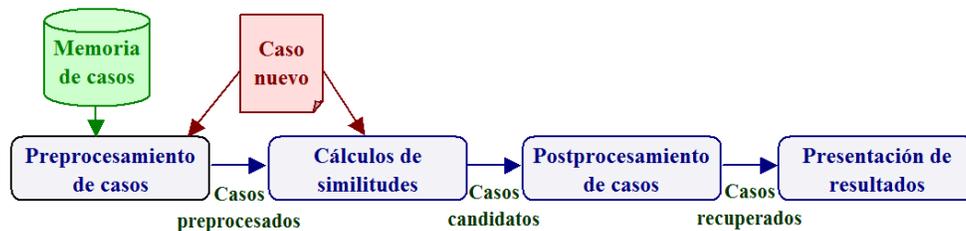


Figura 4: Etapas involucradas en la recuperación de casos arquitectónicos

Durante la primera etapa: **Preprocesamiento de casos** se recorren todos los *casos almacenados en la memoria* analizándolos desde la perspectiva de las restricciones de diseño impuestas por los “stakeholders” en el nuevo caso. Como resultado de esta etapa se obtiene un conjunto de *casos preprocesados*, los cuales son generados a partir de aquellos casos de la memoria que satisfacen las restricciones de diseño obligatorias del nuevo caso. La segunda etapa: **Cálculo de similitudes** está enfocada en los requerimientos de calidad solicitados por los “stakeholders” (documentados mediante escenarios específicos del sistema): se itera sobre todos los casos preprocesados obtenidos en la etapa anterior y se calcula la similitud entre cada uno de ellos y el nuevo caso analizando la similitud que existe entre los escenarios específicos del sistema de ambos casos (nuevo y candidato). Como resultado de esta etapa se obtiene un conjunto de *casos*

candidatos. En la tercera etapa: **Postprocesamiento de casos** se recorren todos los casos candidatos con el objetivo de analizar las decisiones documentadas en ellos para determinar cuán recomendables son las estrategias de diseño que serían propuestas a los arquitectos de software para dar solución al caso nuevo. Los casos candidatos se transforman así en *casos recuperados*. Finalmente, en la cuarta etapa: **Presentación de resultados** se exhiben los resultados obtenidos recordando y recomendando, a partir de cada caso recuperado, estrategias de diseño a los arquitectos de software que participan del diseño de la arquitectura del nuevo caso.

5.3. Reuso, revisión y retención

Como resultado de la actividad de recuperación, los casos obtenidos de la memoria son informados a los arquitectos de software involucrados en el diseño arquitectónico del nuevo sistema. A continuación, y como parte de la técnica de Razonamiento Basado en Casos propuesta, los arquitectos deben analizar las soluciones presentadas y seleccionar una para su *reuso*. La solución seleccionada se constituye como la base para la definición de la nueva arquitectura, no es en sí misma una arquitectura completamente definida, por lo que los arquitectos deberían trabajar sobre ella. Cabe destacar que con la aplicación de la técnica de Razonamiento Basado en Casos propuesta en esta Tesis, no se intenta desplazar o reemplazar al arquitecto ni automatizar por completo su trabajo, sino que el objetivo es facilitar su labor y la de todos aquellos que comparten la memoria de casos, recordándoles experiencias pasadas que pueden ser de utilidad para la construcción del nuevo sistema.

Para que el caso que describe una arquitectura sea retenido en la memoria de casos debe ser previamente *revisado* y, si se detectan errores o problemas, debe ser corregido. Las evaluaciones y verificaciones deben ser llevadas a cabo desde dos puntos de vista:

- desde el punto de vista de diseño ingenieril, efectuando las acciones necesarias para garantizar que la arquitectura diseñada permite satisfacer los intereses de los “stakeholders”, y
- desde el punto de vista de la técnica propuesta, efectuando validaciones para determinar si la arquitectura documentada cumple con los requisitos necesarios para poder constituirse como un caso arquitectónico disponible para ser utilizado como parte de la memoria de casos en un nuevo ciclo de la técnica de Razonamiento Basado en Casos

Finalizada la revisión, el caso puede ser *retenido* en la memoria de casos. De esta manera, estará disponible ante la llegada de un nuevo caso que implique una nueva ejecución del ciclo de Razonamiento Basado en Casos.

6. Utilización y aplicación de la propuesta

La propuesta fue evaluada y analizada desde dos puntos de vista: uno funcional y otro de calidad. Desde el punto de vista funcional, se trabajó sobre un caso de estudio referido al diseño de la arquitectura de un sistema de software en particular con el objetivo de materializar la propuesta en el contexto de casos reales extraídos de la industria. Surge, como producto de la observación del caso de estudio, la demostración de que la propuesta no necesariamente requiere una gran inversión de tiempo por parte de los arquitectos, ya que toda la información obtenida de la etapa de análisis de un ciclo de vida de desarrollo de software (restricciones de diseño y escenarios específicos del sistema) puede ser documentada por los analistas de software. De esta forma, los arquitectos sólo deben volcar en el caso arquitectónico la información que describe a las decisiones de diseño tomadas.

Desde el punto de vista de calidad, se analizó la aplicación de la técnica propuesta desde uno de los aspectos de calidad de software: “performance”. ¿Cuánto tiempo debe esperar el arquitecto de software para contar con la propuesta de estrategias a aplicar una vez que ha iniciado la consulta?. Para analizar el tiempo de procesamiento necesario para dar respuesta a una consulta se llevaron a cabo pruebas de performance con distintos lotes de datos. Para ejecutar estas pruebas se

implementó una aplicación que trabajaba con datos aleatorios y permitía efectuar las mediciones necesarias para conocer el tiempo de procesamiento, tomando como parámetros de entrada la cantidad de casos, la cantidad de escenarios específicos del sistema y la cantidad de restricciones.

Los parámetros de entrada fueron establecidos de manera de poder combinarlos para ejecutar varias pruebas considerando una cantidad de casos en el rango de [20, 200], de escenarios específicos del sistema en el rango de [20, 500] y restricciones de diseño en el rango de [5 y 75].

Como resultado del estudio llevado a cabo se pudo concluir que si bien el tiempo promedio de espera aumenta en casos con muchos escenarios específicos del sistema, o cuando la memoria de casos es muy grande, el valor obtenido está dentro de un rango de valores razonable que no impactaría en el trabajo diario de los arquitectos.

También se trabajó en una herramienta, llamada RADS, que es un prototipo desarrollado para ofrecer a los arquitectos un soporte computacional que formaliza y materializa la propuesta realizada en esta Tesis. Dicha herramienta, implementada en lenguaje JAVA como un complemento de Eclipse, permite la captura de escenarios generales, estrategias de diseño, de los intereses planteados por los “stakeholders” de un sistema, y las decisiones tomadas por los arquitectos de software. Ofreciendo además la posibilidad de recuperar las estrategias de diseño aplicadas en casos previos para emplearlas en nuevos sistemas.

7. Aportes de la Tesis.

Los principales aportes de la Tesis se describen a continuación:

- Definición de un modelo que identifica los conceptos necesarios para representar arquitecturas de software como casos en el contexto de una técnica de Razonamiento Basado en Casos.
- Definición de una metodología que especifica las etapas necesarias para poder aplicar la técnica de inteligencia artificial de Razonamiento Basado en Casos a un problema de ingeniería de software como lo es el diseño arquitectónico.
- Aplicación de la técnica de Razonamiento Basado en Casos utilizando casos modelados mediante el paradigma orientado a objetos cuya estructura es compleja.
- Definición de una estrategia de recuperación compleja.
- Definición de una medida de similitud, desde la perspectiva de requerimientos no funcionales, que permite comparar dos casos arquitectónicos, y por consiguiente, dos arquitecturas de software que hayan sido representadas mediante casos.
- Organización de la información relacionada a las arquitecturas diseñadas en una memoria de casos compartida y de libre acceso que facilita la cooperación entre los arquitectos de software e impulsa la colaboración al establecer un repositorio común para capturar experiencias de diseño.

8. Futuras líneas de investigación.

Los trabajos futuros a partir de esta propuesta se enmarcan en cuatro grandes líneas, las cuales se describen a continuación:

- Seguimiento de las decisiones de diseño: Se considera, en esta línea de investigación, la utilidad de contar con información relacionada con la utilización de las estrategias de diseño reutilizadas como producto del empleo de la propuesta presentada en esta tesis.
- Enriquecimiento de la información suministrada acerca de los estilos arquitectónicos: Esta línea de investigación está impulsada por la necesidad de brindar al arquitecto de software información útil y válida que le permita tomar mejores decisiones durante el diseño arquitectónico.
- Migración de diseños arquitectónicos existentes: En el momento en que una organización decide utilizar la propuesta presentada en esta Tesis es probable que sus arquitectos ya hayan diseñado varias arquitecturas. También es probable que la descripción de dichas arquitecturas y la documentación del razonamiento subyacente a ellas esté en distintos medios, digitales o físicos, como ser archivos de texto, diagramas UML, notas, correos electrónicos, etc. La incorporación a la memoria de casos de las experiencias de diseño arquitectónico vividas previamente permite

comenzar a utilizar la propuesta contando con casos disponibles plausibles de ser recuperados ante una consulta puntual.

- Perfeccionamiento de la herramienta RADS: Esta línea de trabajo tiene como objetivo poder llevar a cabo las modificaciones e implementaciones necesarias para la integración de RADS con otras herramientas de modelado UML empleadas por los arquitectos de software de la industria.

9. Referencias

- [1] L. Bass, P. Clements, and R. Kazman. *Software Architecture in Practice, Second Edition*. Addison-Wesley, 2003.
- [2] P. Clements, F. Bachmann, L. Bass, D. Garlan, J. Ivers, R. Little, P. Merson, Robert Nord, and Judith Stafford. *Documenting Software Architectures: Views and Beyond (2nd Edition)*. Addison-Wesley, 2010
- [3] P. Kruchten. Mommy, where do software architecture come from? *1st International Workshop on Architectures for Software Systems (IWASSI)*, pages 198–205, 1995.
- [4] R. N. Taylor, N. Medvidovic, and E. M. Dashofy. *Software Architecture: Foundations, Theory, and Practice*. Wiley, 2010
- [5] Ch. Hofmeister, P. Kruchten, R. L. Nord, H. Obbink, A. Ran, and P. America. A general model of software architecture design derived from five industrial approaches. *Journal of Systems and Software*, 80(1):106–126, 2007.
- [6] T. B. Bollinger and S. L. Pfleeger. The economics of reuse: Issues and alternatives. In GA Atlanta, editor, *Proceedings of the Eighth Annual National Conference on Ada Technology*, pages 436–447, 1990.
- [7] D. Gentner. Similarity and analogical reasoning. chapter The Mechanisms of Analogical Learning, pages 197–241. Cambridge University Press, 1989.
- [8] H. Gust, U. Krumnack, K. U. Kuhnberger, and A. Schwering. Analogical reasoning: A core of cognition. *Zeitschrift für Künstliche Intelligenz (KI), Themenheft KI und Kognition*, (1):8–12, 2008.
- [9] M. C. Carignano, S. Gonnet, and H. Leone. Reasoning and Reuse in Software Architecture Design: Practices in the Argentine Industry. *SADIO Electronic Journal of Informatics and Operations Research*, 12(1), September 2013.
- [10] L. Bass, P. Clements, and R. Kazman. *Software Architecture in Practice, Third Edition*. Addison-Wesley, 2013.
- [11] P. Eeles and P. Cripps. *The Process of Software Architecting*. Pearson Education, 2009.
- [12] G. Vazquez, J. A. Diaz Pace, and M. Campo. Reusing design experiences to materialize software architectures into object-oriented designs. *Information Sciences*, 259:396–411, February 2014.
- [13] P. Gomes and A. Leitão. A tool for management and reuse of software design knowledge. In Steffen Staab and Vojtech Svátek, editors, *Managing Knowledge in a World of Networks*, volume 4248 of *Lecture Notes in Computer Science*, pages 381–388. Springer Berlin / Heidelberg, 2006.
- [14] B. López. *Case-Based Reasoning: A Concise Introduction*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2013.
- [15] A. Aamodt and E. Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications. IOS Press*, 7(1):39–59, 1994.
- [16] A. Tang, M. Ali Babar, I. Gorton, and J. Han. A survey of architecture design rationale. *The Journal of Systems and Software, Elsevier*, 79:1792–1804, 2006.
- [17] J. Kolodner. An introduction to case-based reasoning. *Artificial Intelligence Review*, 6:3–34, 1992.
- [18] I. Watson and F. Marir. Case-based reasoning: A review. *The Knowledge Engineering Review*, 9(4):327–354, 1994.
- [19] R. Bergmann, K-D. Althoff, S. Breen, M. Göker, M. Manago, R. Traphöner, and S. Wess. Case-based reasoning approaches. In *Developing Industrial Case-Based Reasoning Applications*, volume 1612 of *Lecture Notes in Computer Science*, pages 21–34. Springer Berlin Heidelberg, 2003.
- [21] OMG. Object constraint language, version 2.2. <http://www.omg.org/spec/OCL/2.2>, 2010.
- [20] OMG. Unified modeling language, version 2.3. <http://www.omg.org/spec/UML/2.3>, 2010.
- [22] International Organization for Standardization and International Electrotechnical Commission. Systems and software engineering - architecture description (ISO/IEC/IEEE 42010). edition 01-12-2011. 2011.
- [23] S. Pfleeger and J. Atlee. *Software Engineering: Theory and Practice*. Prentice Hall, 2009.
- [24] L. Bass, J Berget, P. Clements, P Merson, I Ozkaya, and R Sangwan. A comparison of requirements specification methods from a software architecture perspective. Technical Report Tech. Rep. CMU/SEI-2006-TR-013, SEI, 2006.
- [25] L. Bass, M. Klein, and G. Moreno. Applicability of general scenarios to the architecture tradeoff analysis method. Technical Report CMU/SEI-2001-TR-014, SEI, 2001.
- [26] F. Bachmann, L. Bass, and R. Nord. Modifiability tactics. Technical Report CMU/SEI-2007-TR-002, SEI, 2007.
- [27] F. Bachmann, L. Bass, and M. Klein. Illuminating the fundamental contributors to software architecture quality. Technical Report CMU/SEI-2002-TR-025, SEI, 2002.
- [28] R. Prietro-Díaz and P. Freeman. Classifying software for reusability. *IEEE Software*, 4(1):6, 1987.

Marcos Argumentativos Etiquetados

Maximiliano C. D. Budán

Directores: **Guillermo R. Simari** y **Rosanna N. Costaguta**

Fecha de exposición: 16/12/2015

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,
Departamento de Matemáticas, Universidad Nacional de Santiago del Estero,
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina
Tel: (0291) 459-5135 / Fax: (0291) 459-5136
e-mail: mcdb@cs.uns.edu.ar

Resumen

El área de la representación del conocimiento y el razonamiento rebatible en Inteligencia Artificial se especializa en modelar el proceso de razonamiento humano de manera tal de establecer qué conclusiones son aceptables en un contexto de desacuerdo. En términos generales, las teorías de la argumentación se ocupan de analizar las interacciones entre los argumentos que están a favor o en contra de una determinada conclusión, para finalmente establecer su aceptabilidad.

El objetivo principal del presente trabajo es expandir la capacidad de representación de los marcos argumentativos permitiendo representar las características especiales de los argumentos, y analizar cómo éstas se ven afectadas por las relaciones de soporte, agregación y ataque que se establecen entre los argumentos de un modelo que representa una determinada discusión argumentativa. Para ello, añadiremos un meta-nivel de información a los argumentos en la forma de etiquetas extendiendo así sus capacidades de representación, y brindaremos las herramientas necesarias para propagar y combinar las etiquetas en el dominio de la argumentación. Finalmente, utilizaremos la información proporcionada por las etiquetas para optimizar el proceso de aceptabilidad de los argumentos y brindar así resultados más refinados.

1. Introducción

El estudio de la teoría de la argumentación constituye un tema que desde la Grecia clásica ha atraído la atención de filósofos e investigadores relacionados a diferentes áreas de estudio. Este interés se debe, sin duda, a que la argumentación está presente en diversos aspectos de nuestras vidas, ya sea en aquellas situaciones cotidianas más sencillas o en aquellos debates más complejos. Descripto de una manera simple, la *argumentación* es un discurso expositivo que tiene como finalidad la intención de persuadir o convencer a alguien de dar apoyo a una “*postura*” o “*tesis*”, con la intención de ganar su asentimiento o adhesión a la causa argumentada. La argumentación tiene tres objetivos principales: *identificar*, *analizar* y *evaluar* argumentos que están a favor o en contra de una tesis. Es

usual utilizar el término *argumento* para referirse a la entrega de razones que apoyen o refuten una tesis, la cual debe ser cuestionable o abierta a duda. En un sentido lógico formal, un argumento está compuesto por tres elementos: una conclusión, un conjunto de premisas que respaldan dicha conclusión, y un mecanismo de inferencia que permite alcanzar la conclusión a partir del conjunto de premisas [2].

La argumentación discute las situaciones problemáticas del mundo real a través de un proceso analítico, denominado *proceso argumentativo*, el cual puede presentarse como un juego en donde intervienen tres entidades: un *proponente*, un *oponente* y un *juez* o *árbitro* o *jurado*. Por un lado, el *proponente* desempeña la función de introducir la tesis y brindar los argumentos que la soportan. Por otro lado, el rol del *oponente* es el de interponerse al propósito del proponente, y para ello, ofrece argumentos que contradicen o refutan la tesis, o los argumentos introducidos por el proponente. Los argumentos del proponente se denominan argumentos *pro* y los del oponente se denominan *op*. Un argumento ofrecido para batir a otro argumento se denomina *contra-argumento*. Así, el proceso argumentativo comienza cuando el proponente introduce una tesis junto a los argumentos *pro* que soportan la misma, luego es el turno del oponente quien ofrece sus contra-argumentos. En este momento, el proponente se transforma en oponente de su contrincante y ofrece contra-argumentos para los argumentos introducidos por su contrincante. El proceso continúa de esta manera hasta agotarse. Una vez que se tienen presentes todos los argumentos *op* y *pro*, entra en juego el papel del *arbitro* o *jurado* o *audiencia*, quien determina cual de estos argumentos serán aceptados, y pasarán a ser considerados como creencias o verdades. Finalmente, el conjunto de creencias que un agente o sistema inteligente posee es usado con diferentes propósitos, como ser solucionar alguna situación problemática del mundo real de una manera eficiente u óptima.

Desde los años 80, la Inteligencia Artificial (IA) ha buscado producir avances frente al desafío de imitar el mecanismo de razonamiento que los seres humanos generalmente empleamos para debatir acerca de algún tema específico, ya sea con otros seres humanos o internamente con uno mismo, con el propósito de especificar una base de creencias que puede ser utilizada para razonar de manera inteligente frente a determinadas situaciones problemáticas. En este sentido, para lograr este comportamiento, es necesario identificar la situación problemática, representar el conocimiento disponible, y tratar de razonar sobre esta representación para alcanzar una óptima solución en base a la información disponible. Como se explicó anteriormente, es usual que el conocimiento relacionado a la situación planteada sea inconsistente. Es por ello que, se necesitan técnicas de representación y razonamiento que permitan abordar estos problemas. Existen dos visiones que se enfocan a solucionar o tratar el problema del conocimiento inconsistente, los cuales son: *restaurar* la consistencia, objeto de estudio de la *Revisión de Creencias*; o *razonar* en un modelo que contiene inconsistencia, construyendo y evaluando argumentos que soportan conclusiones contradictorias, objeto de estudio de la *Argumentación Rebatible* [16].

La argumentación rebatible es una formalización del razonamiento rebatible [17] donde se pone especial énfasis en la noción de argumento. En particular, un argumento para una conclusión *C* constituye una pieza de razonamiento tentativa que un agente inteligente está dispuesto a aceptar para explicar *C*. Si el agente adquiere luego nueva información, la conclusión *C* junto al argumento que la soporta podrían quedar invalidados. En un sistema argumentativo rebatible la validez de una conclusión *C* será garantizada, cuando exista un argumento que brinde una justificación válida para *C*. Este proceso involucra la construcción de un argumento *A*, para *C*, que no se encuentre derrotado. En este sentido, para verificar si el argumento *A* está derrotado, se construyen contra-argumentos que son posibles derrotadores de *A*. Como estos derrotadores son argumentos, se debe verificar que no estén a su vez derrotados, y así siguiendo. De esta manera se modela el proceso de razonamiento en el cual se producen y se evalúan argumentos a favor y en contra

de una conclusión para verificar la garantía de dicha conclusión [17]. Los formalismos argumentativos crean modelos argumentativos para representar y analizar las diferentes situaciones problemáticas del mundo real. Cada uno de estos modelos posee diferentes niveles de abstracción, dependiendo del dominio de estudio para el que dichos modelos son creados. Actualmente, los formalismos basados en la argumentación rebatible han sido aplicados con éxito a diferentes problemas de la IA, tales como negociación, razonamiento legal, sistemas de recomendación y de conciliación de ontologías, entre otros.

2. Motivaciones

Los formalismos argumentativos clásicos brindan la posibilidad de crear modelos que permiten representar el conocimiento en forma de argumentos, analizarlos, y evaluarlos para determinar cuáles son aptos para respaldar la toma de decisiones o realizar acciones de manera inteligente. En estos formalismos, el análisis y la evaluación de los argumentos se realiza en base a dos factores: las propiedades que determinan la solidez lógica de un argumento, y las relaciones definidas entre argumentos. Sin embargo, el mundo real es incierto e impreciso, es por esto que la realidad no puede estudiarse en términos absolutos con técnicas aplicables a situaciones ciertas, que en la búsqueda de la precisión, intentan ajustar el mundo real a modelos matemáticos rígidos y estáticos perdiendo con ello información valiosa. Por ello, según L. Zadeh, al abordar el estudio de fenómenos y a medida que la complejidad de un sistema aumenta, nuestra capacidad de realizar formulaciones precisas y significativas sobre su comportamiento disminuye hasta que se alcanza un umbral por debajo del cual la precisión y la relevancia se convierten en características mutuamente excluyentes [19]. En base a este análisis, los formalismos argumentativos deberían poseer la capacidad de representar no sólo las propiedades relacionadas a la solidez lógica de los argumentos, sino también otras propiedades dependientes del dominio de aplicación que proporcionen modelos argumentativos más representativos de la realidad con el objetivo de optimizar los procesos de razonamiento que se llevan a cabo sobre dichos modelos. Esta intuición sugiere que sería beneficioso aumentar la capacidad de representación de las estructuras argumentales para permitir modelar las cualidades especiales de los argumentos. La noción de valorar y ponderar argumentos, fue inicialmente introducida por Bench-Capon en [1], y ha comenzado a cobrar cada vez más importancia en los últimos años. Bench-Capon asocia su pensamiento al trabajo de jurisprudencia propuesto por Perelman [13], el cual puede ser tomado como fuente de ejemplos, donde las valoraciones asociadas a los argumentos se muestra como algo natural en el razonamiento humano.

Los procesos argumentativos que incorporan valoraciones asociadas a los argumentos pueden ser estudiados en dos etapas: la determinación de las valoraciones asociadas a los argumentos, y la selección de los argumentos aceptados. La valoración de un argumento puede ser obtenida independientemente de las interacciones definidas con otros argumentos, o puede ser dependiente de las relaciones que un argumento posee con los demás argumentos del modelo, tales como soporte, ataque, agregación y conflicto. En cuanto a la selección del conjunto de los argumentos que se encuentran aceptados, es posible realizar un análisis en dos direcciones: la aceptabilidad individual y la aceptabilidad colectiva. En el primer caso, la aceptabilidad de un argumento depende enteramente de sus atributos. En el segundo caso, la aceptabilidad de un conjunto de argumentos depende del cumplimiento de ciertas propiedades predefinidas.

En los últimos años, surgieron diversos formalismos modelando distintos aspectos del mundo real bajo distintos fines u objetivos. Por ejemplo, Cayrol y Lagasque-Schieh en [7] propusieron un marco argumentativo en donde persiguen como propósito introducir gradualidad en la selección de los mejores argumentos, presentando así diferentes niveles de aceptabilidad. En este caso, las valoraciones asociadas a los argumentos dependen de la re-

lación de ataque que se producen entre los mismos, es decir, a mayor número de atacantes menor es la fuerza del argumento atacado. Por otro lado, Joao Leite and Joao Martins en [11] presentan una extensión del marco argumentativo de Dung donde se incorporan votos sociales asociados a cada argumento; así, se añade el aspecto social a los modelos argumentativos representando el contexto en el cual se lleva a cabo el debate. De esta manera, es posible realizar votos a favor y en contra de cada argumento, afectando de manera positiva o negativa la fortaleza de cada uno de ellos. En este formalismo, todos los ataques definidos sobre los argumentos tienen el mismo impacto, es decir, la fuerza del ataque no tiene en cuenta los diferentes niveles de experticia de los votantes. En base a esta postura, Egilmez *et al.* en [8] presentan una extensión, en donde es posible asignar votos a los ataques, brindando la posibilidad de reflejar variaciones en la fuerza de los ataques producidos entre argumentos. Por otro lado, Pollock en [14] argumenta que en la mayoría de las semánticas para el razonamiento rebatible no se tienen en cuenta el hecho de que algunos argumentos son mejores que otros, ofreciendo así diferentes niveles de apoyo para sus correspondientes conclusiones.

En bases a estas líneas de investigación podemos concluir que los formalismos argumentativos necesitan contar con la capacidad de representar las características de los argumentos que son dependientes del dominio de aplicación, con el objetivo de crear modelos argumentativos más representativos del mundo real. En esta dirección, sería posible lograr una satisfactoria integración de la argumentación en los diferentes dominios de aplicación, tales como agentes autónomos en sistemas de soporte a las decisiones, búsqueda inteligente en la web, administración del conocimiento, sistemas de recomendación, y otros dominios de similar importancia.

3. Contribuciones y resultados de la tesis

El objetivo principal de esta línea de investigación es expandir la capacidad de representación de los formalismos argumentativos. En términos generales, en este trabajo, se presentan dos formalismos, cada uno de ellos con un cierto nivel de abstracción, que permiten considerar la meta-información dependiente del dominio de aplicación dentro del proceso de razonamiento argumentativo. Esta meta-información estará asociada a los argumentos tomando la forma de etiquetas, incrementando así su capacidad de representación. Dichas etiquetas pueden ser afectadas por las relaciones existentes entre los argumentos del modelo. Por esta razón, se define una estructura algebraica, llamada álgebra de etiquetas argumentales, que permite la combinación y propagación de la información asociada a los argumentos en el dominio de la argumentación. Así, la introducción de las etiquetas nos brinda la posibilidad de representar las características asociadas a los argumentos, tales como grado de incertidumbre, grado de confiabilidad, valores posibilísticos, valores probabilísticos, medidas de fuerza, o cualquier propiedad relevante, proporcionando la herramienta para refinar el proceso por el cual se determina la aceptabilidad de los argumentos, y la calidad de garantía de una determinada conclusión.

A continuación introduciremos brevemente los tres aportes principales que se realizan en esta tesis. En primer lugar, describiremos las estructuras y operaciones algebraicas que se utilizarán para representar y propagar las características especiales de los argumentos. Luego, presentaremos los formalismos argumentativos etiquetados (o valuados) que se desarrollaron con el objetivo de expandir la capacidad de representación de los formalismos argumentativos actuales, ampliando el alcance de los procesos argumentativos que es posible modelar a través de las teorías argumentativas. Los resultados de las contribuciones expuestas en esta tesis están incluidos en los artículos *Modelling time and reliability in structured argumentation frameworks* publicado en *The Conference on the Principles of Knowledge Representation and Reasoning* [6], *Modeling time and valuation in structured*

argumentation frameworks publicado en *The Journal Information Sciences An aif-based labeled argumentation framework* publicado en *The Conference on Foundations of Information and Knowledge Systems* [5], *A labeled argumentation framework* publicado en *The Journal of Applied Logic* [4], entre otras.

Álgebra de Etiquetas Argumentales

En esta tesis se propuso la formalización de un álgebra que consiste de un conjunto de etiquetas equipadas con una colección de operadores que permiten combinar y propagar estas etiquetas según las relaciones que se producen entre los argumentos, tales como soporte, conflicto y agregación. Así, se introduce el uso de etiquetas como una herramienta para ayudar a la caracterización y evaluación de los argumentos. Para ser de utilidad, estas etiquetas deben contener información distintiva de los argumentos, y la manera en la que éstos interactúan en el dominio de la argumentación. Una forma natural de representar esta información es la utilización de una escala que mida las características distintivas de los argumentos, como ser el grado de confiabilidad del mismo en base a la confiabilidad que se tiene de las fuentes que los proporcionan, la relevancia de la información que proporciona un argumento dependiendo de los puntos de análisis de un usuario, entre otros. En concreto, el álgebra de etiquetas se basa en un conjunto ordenado que permite la comparación de las *etiquetas argumentales*, y este conjunto se caracteriza de la forma más abstracta posible para adaptarse a los distintos requisitos que poseen las diferentes aplicaciones del dominio de la argumentación. Formalmente:

Definición 1 *Un Álgebra de Etiquetas Argumentales esta compuesta por una 7-tupla que tiene la forma $A = \langle A, \leq, \odot, \oplus, \ominus, \top, \perp \rangle$, donde:*

- A es un conjunto de etiquetas argumentales, llamado dominio de etiquetas;
- \top y \perp son dos elementos distintivos de A , tal que \top es el mayor elemento, mientras que \perp es el menos de ellos;
- \leq es una relación de orden parcial sobre el conjunto de etiquetas argumentales A , donde \leq es reflexiva, antisimétrica y transitiva;
- $\odot : A \times A \rightarrow A$ es denominado el operador de soporte *satisfaciendo las propiedades conmutativa, asociativa y monótona;*
- $\oplus : A \times A \rightarrow A$ es denominado el operador de agregación *satisfaciendo las propiedades conmutativa, asociativa y monótona;* y
- $\ominus : A \times A \rightarrow A$ es denominado el operador de conflicto *satisfaciendo las propiedades modulativa, uniforme y monótona.*

El operador de soporte, denotado como “ \odot ”, es usado para determinar la valuación de un argumento basado en las valuaciones de los argumentos que lo soportan. Esta dependencia debe ser invariante del orden con el cual se consideran los argumentos de apoyo, y por lo tanto, la operación de soporte debe ser conmutativa y asociativa. Asimismo, la propiedad de monotonocidad del operador “ \odot ” es justificada por el *principio del eslabón más débil*, puesto que si un argumento es soportado por un conjunto de argumentos, su valuación debe ser menor o igual a la menor valuación de los argumentos que lo soportan. Por otro lado, el operador de agregación, denotado como “ \oplus ”, es usado para determinar la valuación de una conclusión basada en las valuaciones de las diferentes fuentes “*independientes*” que la soportan. La agregación de argumentos fue cuidadosamente motivada dentro del dominio de la argumentación basándose en la siguiente intuición: mientras mayor sean las razones para soportar una determinada conclusión, mayor será la credibilidad de la misma [18, 15, 10]. La forma más natural de realizar la agregación de argumentos es

efectuando una adición de las valuaciones de las fuentes que soportan dicha conclusión, o aplicar alguna operación que sea una variación de la adición, dependiendo de la clase de información que se desea operar. El operador “ \oplus ” tiene algunas de las propiedades de la adición. De esta manera, “ \oplus ” debe ser conmutativa y asociativa, con \perp como elemento neutro tal que una fuente con dicha valuación no represente ningún significado en el momento de obtener la valuación agregada final. Finalmente, el operador de conflicto, denotado como “ \ominus ”, actúa como una clase de sustracción con respecto a la adición de los números reales en relación al operador de agregación “ \oplus ”. En un sentido general, el operador de conflicto determina las valuaciones de un argumento al momento de considerar las valuaciones asociadas a las razones contrapuestas. Así, un argumento puede ser debilitado, y en algunos casos derrotado, según la fuerza de sus contra-argumentos. Al caracterizar el operador de conflicto se establecieron todas las condiciones necesarias para garantizar una coherente resolución de los conflictos dentro del dominio de la argumentación. En este sentido, el operador satisface las propiedades: modulativa, uniforme y monótona.

Existen diferentes ejemplos de instanciación para el álgebra de etiquetas argumentales, y es importante determinar cual de ellas es la más apropiada para modelar los diferentes casos del mundo real. Esta selección es principalmente metodológica, por ello, es necesario analizar la semántica del dominio de diferentes maneras posibles, como ser: diseñar experimentos usando ejemplos donde se conoce la conclusión deseada, realizar pruebas utilizando la evaluación cognitiva de sujetos humanos aproximando sus evaluaciones a las valoraciones de argumentos obtenidos después de computar sus interacciones, entre otros. [?].

Marco Argumentativo Estructurado General Etiquetado

Se considera una sucesión de formalismos argumentativos que van evolucionando en nivel de detalle y capacidad de representación del mundo real, partiendo del marco argumentativo abstracto propuesto por Dung hasta llegar a un marco argumentativo estructurado general etiquetado (*Figura 1*).

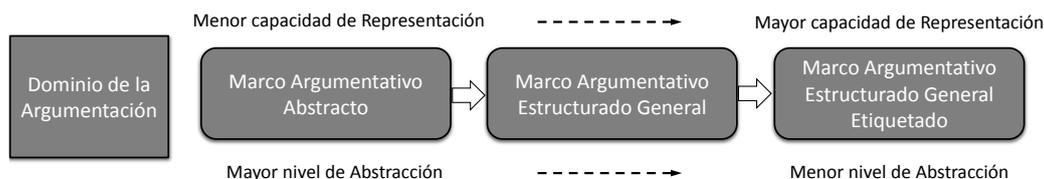


Figura 1: Marco argumentativo abstracto / Estructurado / Etiquetado

En primer lugar, se presenta un formalismo denominado *Marco Argumentativo Estructurado* (*GeSAF*, por su sigla en inglés), que permite representar la estructura interna de los argumentos a través de estructuras argumentales, y establecer diferentes tipos de relaciones entre ellas, tales como conflicto, preferencia y derrota. De esta manera, *GeSAF* mantiene un cierto grado de abstracción, permitiendo describir la estructura interna de los argumentos teniendo en cuenta la información que forma parte de sus estructuras, tales como pasos de razonamiento, suposiciones y evidencias. Existen dos razones que justifican el desarrollo de un formalismo que permite pasar de un marco argumentativo abstracto a un marco argumentativo estructurado. Por una parte, su introducción permite una generalización de diferentes sistemas argumentativos estructurados, tales como ABA [3], ASPIC+ [12], o DeLP [9], sin tener que comprometerse específicamente a uno de ellos; en segundo lugar, se pueden generalizar las nociones de aceptabilidad que capturan el proceso de aceptabilidad, posibilitando la adaptación de dicho proceso a los cambios del dominio de la aplicación, analizando así un mismo modelo argumentativo aplicando

diferentes semánticas. Luego, se introduce el desarrollo de un marco argumentativo estructurado general etiquetado (*L-GeSAF*, por su sigla en inglés), en el cual se incorpora las nociones y estructuras formales necesarias para asociar meta-información en forma de etiquetas a cada uno de los argumentos, tales como su grado de confiabilidad, votos sociales, entre otros. Por lo general, esta información no se encuentra asociada directamente a los argumentos sino que está relacionada a las piezas básicas del conocimiento a partir de las cuales son construidos. En este sentido, sería interesante determinar las cualidades de los argumentos que intervienen en una discusión argumentativa en base a las cualidades asociadas a las piezas de conocimiento que integran la misma con la intención de tener una noción de su fuerza colectiva. Por ello, usaremos un álgebra de etiquetas argumentales como la herramienta para posibilitar su correcta combinación y propagación dentro del dominio de la argumentación. En *L-GeSAF*, la información asociada a los argumentos puede ser usada de diversas formas dependiendo de los fines que se desea alcanzar, como ser proporcionar información adicional acerca de la aceptabilidad de los argumentos (por ejemplo, el nivel de confiabilidad de los argumentos, el valor posibilístico o probabilístico asociado a los argumentos, entre otros), establecer la calidad de garantía de una conclusión en base a la agregación de las cualidades de los argumentos que la soportan, y definir un umbral de garantía que establece las condiciones que una determinada conclusión debe satisfacer para ser considerada válida.

Marco Argumentativo Etiquetado

Se presenta un formalismo argumentativo etiquetado que permite representar las características de los argumentos del modelo que representa una situación problemática del mundo real, partiendo como base de un modelo argumentativo creado por una ontología argumentativa llamada *Formato Estándar para el Intercambio de Argumentos* (Figura 2).

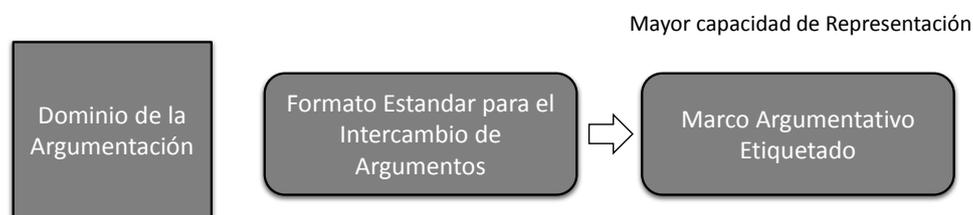


Figura 2: Marco Argumentativo Etiquetado

De esta manera, se introduce un formalismo argumentativo para facilitar la representación del conocimiento conocido como *Formato Estándar para el Intercambio de Argumentos* (*AIF*, por su sigla en inglés), compuesto por un conjunto de conceptos de alto nivel relacionados con el dominio de la argumentación. En *AIF*, se introduce una ontología para expresar las relaciones entre estructuras argumentales con el objetivo de proporcionar un puente entre modelos lingüísticos, lógicos y formales de argumentación y razonamiento. Es importante notar que la ontología de *AIF* está pensada puramente como un lenguaje para expresar argumentos y representar las relaciones existentes entre ellos. Luego, se propone la construcción de un formalismo llamado *marco argumentativo etiquetado* (*LAF*, por su sigla en inglés), en donde se combina las capacidades de representación del conocimiento proporcionadas por *AIF* con el procesamiento de meta-información definido por el álgebra de las etiquetas argumentales. Este marco argumentativo nos permitirá representar argumentos teniendo en cuenta su estructura interna, modelar las diferentes relaciones entre argumentos, y adjuntar a los argumentos sus características especiales a través de etiquetas argumentales. Las interacciones entre argumentos tales como soporte, conflicto

y agregación, tienen asociadas operaciones en el álgebra de etiquetas argumentales permitiendo plasmar el comportamiento del conocimiento en el dominio de la argumentación. Finalmente, se utilizará la información proporcionada por dichas etiquetas para alcanzar diferentes propósitos, tales como proporcionar información adicional acerca de la aceptabilidad de los argumentos, establecer diferentes grados de aceptabilidad en base a las cualidades de los argumentos, definir un umbral de calidad en donde se establezcan las condiciones necesarias para que un argumento sea considerado lo suficientemente fuertes como para ser aceptado, brindar la posibilidad de analizar las posibles soluciones a un modelo argumentativo que representa una determinada situación problemática estableciendo los escenarios que optimicen la justificación de una determinada conclusión.

4. Conclusiones

En la actualidad, el estudio de la argumentación ha recobrado vigencia debido a la gran influencia que los medios de comunicación tienen sobre la sociedad. Esta influencia se manifiesta en el planteamiento de estrategias argumentativas para convencer al público acerca de ciertos valores e ideas. Ejemplo de esto son los discursos argumentativos relacionados con la publicidad o el pensamiento político. Así pues, la principal motivación del estudio de la argumentación consiste en establecer si el razonamiento planteado es verosímil, es decir, si quien es objeto de la argumentación estará dispuesto a aceptarla. En este sentido, los formalismos argumentativos deben poseer la capacidad de representar el discurso argumentativo que plasma una determinada situación problemática del mundo real, y determinar la validez de los argumentos participantes. No obstante, el mundo real es muy complejo e impreciso, es por ello que la realidad no puede estudiarse en términos absolutos. Es decir, la realidad no puede analizarse ni modelarse sin tener en cuenta ciertas características tales como la incertidumbre o la imprecisión de la información que poseemos sobre la situación del mundo real que pretendemos estudiar. Por ello, es correcto pensar que los formalismos argumentativos deberían poseer la capacidad de representar no sólo las propiedades relacionadas a la solidez lógica de los argumentos, sino también otras propiedades dependientes del dominio de aplicación, proporcionando modelos argumentativos más cercanos a la realidad con la intención de optimizar los procesos de razonamiento que se llevan a cabo sobre dichos modelos.

En este trabajo se introdujo el uso de etiquetas como una herramienta para ayudar a la caracterización y evaluación de los argumentos. En particular, definimos un *álgebra de etiquetas argumentales* como una estructura algebraica abstracta, donde se establece el conjunto de operaciones necesarias para manipular las características asociadas a los argumentos. En este sentido, los efectos del soporte, la agregación, y el conflicto entre argumentos se reflejan en las etiquetas argumentales que cada argumento tiene asociada. De esta manera, las etiquetas cumplen la función de informar cómo los argumentos se afectan entre sí. En este sentido, dentro de los formalismos presentados en esta tesis, las etiquetas argumentales que proporcionan información sobre las características de los argumentos son usadas para diversos fines, tales como: (i) evaluar los estados de aceptabilidad asociados a los argumentos, ya que la información proporcionada por las etiquetas argumentales colabora en el proceso de aceptabilidad capturando el comportamiento del conocimiento dentro del modelo argumentativo; (ii) especificar una *relación de preferencia*, ya sea parcial o total, sobre el conjunto de argumentos en base a sus características especiales; (iii) analizar la resolución de conflictos entre argumentos reflejando la disminución bidireccional de las cualidades; (iv) establecer diferentes grados de aceptabilidad usando la información asociada a los argumentos; (v) introducir un *umbral de calidad* con el objetivo de establecer los requerimientos mínimos que un argumento debe satisfacer

para formar parte de la justificación que sustenta una determinada afirmación, decisión o acción; (vi) analizar un conjunto de soluciones que representen las diferentes propagaciones de los atributos asociados a los argumentos que conforman un determinado modelo argumentativo con la intención de identificar aquella solución que optimice las cualidades de una determinada conclusión; y (vii) combinar diferentes características con la intención de proporcionar respuestas más contundentes sobre el estado de una conclusión.

5. Líneas de investigación futuras

A continuación se incluyen algunas líneas de investigación relacionadas a esta tesis sobre las cuales se pretende seguir trabajando:

- En *LAF* la complejidad para resolver el sistema de ecuaciones que modela el comportamiento del conocimiento está asociada a la selección de las operaciones que determinan cómo las etiquetas argumentales serán afectadas por las interrelaciones del conocimiento dentro del grafo argumentativo. La linealidad de un sistema admite ciertas suposiciones matemáticas y aproximaciones que permiten un cálculo más sencillo de los resultados ya que los sistemas no lineales son usualmente difíciles (o imposibles) de tratar, y sus comportamientos con respecto a una variable dada es extremadamente difícil de predecir. En futuros trabajos se pretende profundizar en la caracterización del álgebra de etiquetas argumentales con la intención de establecer las condiciones necesarias que aseguren la creación de un sistema de ecuaciones matemáticamente solubles y computacionalmente tratables.
- En ciertas situaciones del mundo real se necesita la representación de la disponibilidad temporal de los argumentos dentro del dominio de la argumentación con la intención de crear modelos argumentativos dinámicos que contemplen cómo los argumentos son afectados por los eventos del mundo en un momento determinado. En base a esta intuición, definiremos un álgebra de etiquetas temporales que nos permita manipular y propagar la disponibilidad temporal de las piezas de conocimiento que forman parte de la estructura interna de un argumento, creando así un modelo argumentativo dinámico que modele cambios en las relaciones establecidas entre las estructuras argumentales. Asimismo, se propone modelar la variación de los atributos de un argumento en el tiempo. De esta manera, se formalizará el proceso de aceptabilidad de los argumentos teniendo en cuenta las características asociadas a los argumentos, donde éstas poseen una variación o distribución temporal.
- Los sistemas argumentativos basados en reglas (SABR) son formalismos de argumentación en donde el conocimiento de un agente incluye un conjunto de reglas de inferencia a partir de las cuales se pueden construir argumentos a favor o en contra de una afirmación. Estos sistemas son de particular interés en el área de Inteligencia Artificial dado que este tipo de reglas de inferencia permiten representar conocimiento de sentido común, y la construcción de argumentos puede realizarse de manera automática. Como futuro trabajo se implementará un sistema argumentativo valorado estructurado, extendiendo las capacidades de representación de DeLP [9], donde las valuaciones de los argumentos puedan variar en el tiempo, y se analizará cómo éstas variaciones influyen en el proceso de aceptabilidad de los argumentos.

Referencias

- [1] Trevor J. M. Bench-Capon. Value-based argumentation frameworks. In Salem Benferhat and Enrico Giunchiglia, editors, *NMR*, pages 443–454, 2002.

- [2] Philippe Besnard and Anthony Hunter. *Elements of argumentation*, volume 47. MIT press Cambridge, 2008.
- [3] Philippe Besnard and Anthony Hunter. Constructing argument graphs with deductive arguments: a tutorial. *Argument & Computation*, 5(1):5–30, 2014.
- [4] Maximiliano CD Budán, Mauro Gómez Lucero, Ignacio Viglizzo, and Guillermo R Simari. A labeled argumentation framework. *Journal of Applied Logic*, 13(4):534–553, 2015.
- [5] Maximiliano CD Budán, Mauro J Gómez Lucero, and Guillermo Ricardo Simari. An aif-based labeled argumentation framework. In *Foundations of Information and Knowledge Systems*, pages 117–135. Springer, 2014.
- [6] Maximiliano Celmo Budán, Mauro Gómez Lucero, Carlos Iván Chesñevar, and Guillermo Ricardo Simari. Modelling time and reliability in structured argumentation frameworks. In *Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning*, 2012.
- [7] Claudette Cayrol and Marie-Christine Lagasque-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research*, 23:245–297, 2005.
- [8] Sinan Eçilmez, Joao Martins, and Joao Leite. Extending social abstract argumentation with votes on attacks. In *Proceedings of Theory and Applications of Formal Argumentation*, pages 16–31. Springer, 2014.
- [9] Alejandro J García and Guillermo R Simari. Defeasible logic programming: Delp-servers, contextual queries, and explanations for answers. *Argument & Computation*, 5(1):63–88, 2014.
- [10] Mauro J Gómez Lucero, Carlos I Chesñevar, and Guillermo R Simari. Modelling argument accrual in possibilistic defeasible logic programming. In *Proceedings of Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 131–143. Springer, 2009.
- [11] Joao Leite and Joao Martins. Social abstract argumentation. In *Proceedings of International Joint Conferences on Artificial Intelligence*, pages 2287–2292, 2011.
- [12] Sanjay Modgil and Henry Prakken. The aspic+ framework for structured argumentation: a tutorial. *Argument & Computation*, 5(1):31–62, 2014.
- [13] Chaim Perelman. *Justice, Law and Argument*, volume 142 of *Synthese Library*. Reidel, Dordrecht, Holland, 1980.
- [14] John L. Pollock. Defeasible reasoning and degrees of justification. *Argument & Computation*, 1(1):7–22, 2010.
- [15] Henry Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *Proceeding of International Conference on Artificial Intelligence and Law*, pages 85–94. ACM, 2005.
- [16] Iyad Rahwan and Guillermo R. Simari. *Argumentation in Artificial Intelligence*. Springer Verlag, 2009.
- [17] Guillermo R Simari and Ronald P Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial intelligence*, 53(2):125–157, 1992.
- [18] B. Verheij. Accrual of arguments in defeasible argumentation. In *Proceedings of Workshop on Nonmonotonic Reasoning*, pages 217–224, 1995.
- [19] Lotfi A Zadeh, King-Sun Fu, and Kokichi Tanaka. Fuzzy sets and their applications to cognitive and decision processes:. In *Proceedings of Seminar on Fuzzy Sets and Their Applications*. Academic press, 1974.

Consolidación de Ontologías Datalog[±]

Cristhian Ariel David Deagustini

Directores: Guillermo R. Simari y Marcelo A. Falappa

Fecha de exposición: 25 de noviembre de 2015

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina
Tel: (0291) 459-5135 / Fax: (0291) 459-5136
e-mail: cadd@cs.uns.edu.ar

Resumen

En la presente tesis nos enfocamos en el manejo de dos problemas diferentes pero relacionados que suelen aparecer en el conocimiento, especialmente en entornos colaborativos: inconsistencias e incoherencias. Inconsistencia es un problema clásico y ampliamente reconocido en la representación de conocimiento, el cual trae importantes consecuencias para los mecanismos clásicos de inferencia. Incoherencia, por otra parte, ha recibido cada vez más atención desde el surgimiento de lenguajes ontológicos; la misma se relaciona con conflictos en el conjunto de reglas ontológicas que hacen a tales reglas imposibles de satisfacer al mismo tiempo. En este trabajo formalizamos la noción de incoherencia en ontologías Datalog[±], que se encontraba ausente en la literatura, en términos de la satisfacibilidad del conjunto de restricciones en las mismas, y mostramos como bajo ciertas condiciones incoherencia puede llevar a ontologías Datalog[±] inconsistentes.

La contribución principal de este trabajo es el desarrollo de dos operadores noveles para la restauración tanto de la consistencia como la coherencia en ontologías Datalog[±]. Los enfoques propuestos se basan en kernel contraction. En el primero de ellos la restauración se realiza mediante la aplicación de funciones de incisión que seleccionan fórmulas para remoción de los conjuntos incoherentes/inconsistentes mínimos encontrados en las ontologías. Tal operador trata los conflictos mínimos de manera local, sin tener en cuenta la relación (si es que existe) entre los diferentes conflictos mínimos. El otro enfoque, que puede ser visto como un enfoque global, tiene en cuenta tal relación mediante el agrupamiento de conflictos mínimos relacionados en nuevas estructuras llamadas clusters, mediante el uso de una relación de solapamiento.

En esta disertación presentamos construcciones tanto para el enfoque local como el global, junto con las propiedades que se espera que los mismos satisfagan, expresadas a través de postulados. Finalmente, establecemos la relación entre las construcciones y las propiedades mediante el uso de teoremas de representación. Si bien la propuesta esta enfocada en la consolidación de ontologías Datalog[±], estos operadores pueden ser aplicados a otros tipos de lenguajes ontológicos, tales como las Lógicas Descriptivas, haciéndolos aptos para su uso en ambientes colaborativos como la Web Semántica.

1. Introducción

En tiempos donde la colaboración e interacción entre sistemas se ha vuelto moneda corriente es cada vez más frecuente el encontrarse con problemas derivados de tal interacción, apareciendo a su vez desafiantes oportunidades de investigación asociadas a los mismos. En particular, el manejo de información conflictiva es un importante problema que debe ser atacado [7, 8, 12, 2], especialmente cuando se integra conocimiento proveniente de diferentes fuentes [4, 1], o cuando tal conocimiento será explotado por procesos de razonamiento automáticos. El más conocido dentro de los conflictos en información es el de la inconsistencia. Inconsistencia es un concepto clásico dentro de la Representación de Conocimiento y el Razonamiento, y se refiere a teorías tales que no es posible encontrar para ellas un modelo. La importancia del manejo de la inconsistencia surge del problema que la misma acarrea en las relaciones de consecuencia clásicas: las mismas son *explosivas* [12, 7], en el sentido de que cualquier fórmula es consecuencia de una contradicción lógica, pudiendo por lo tanto derivar el lenguaje completo a partir de una base de conocimiento inconsistente, haciendo a las conclusiones obtenidas completamente irrelevantes.

En adición a la inconsistencia, en entornos ontológicos podemos encontrar otro tipo de conflictos, relacionados a un fenómeno conectado en cierta forma con inconsistencia pero a su vez con sus características propias: incoherencia [6, 5, 3, 13, 17, 16], un concepto que ha surgido con el advenimiento de entornos ontológicos como Datalog[±]. Tal fenómeno versa en una característica del conocimiento expresado a través de las reglas que hace que el mismo no pueda ser aplicado sin generar problemas de consistencia (es decir, un conjunto de reglas las cuales no pueden ser aplicadas sin violar *inevitablemente* alguna de las restricciones impuestas al conocimiento, haciéndolas por lo tanto insatisfacibles). Intuitivamente podemos ver a la incoherencia como una inconsistencia latente o potencial; esto es, el conocimiento incoherente no representa una violación de consistencia en sí mismo, pero cuando un conocimiento ontológico (un conjunto de reglas) incoherente es considerado junto con hechos relevantes (hechos que activen las reglas en cuestión) entonces la violación es inevitable.

En la Inteligencia Artificial (IA) muchos trabajos se han enfocado en el manejo de los problemas de inconsistencia (y también, aunque en menor medida, incoherencia) en conocimiento de forma tal que se evite de alguna manera los problemas acarreados por la misma. En la presente tesis presentamos dos métodos noveles para la resolución de todos los conflictos de coherencia y consistencia en ontologías Datalog[±]. Tales métodos se centran en modificar de alguna forma el conocimiento inconsistente de forma tal que la consistencia en el mismo sea recuperada, para luego aplicar relaciones de consecuencia clásicas sobre el conocimiento consistente obtenido. De esta forma, si bien la base de conocimiento original es inconsistente y por lo tanto capaz de derivar todo el lenguaje, al utilizar una modificación consistente de la misma se pueden aplicar relaciones de inferencia clásicas de manera segura; esto es conocido en la literatura como *consolidación de conocimiento* [11, 10].

Los operadores propuestos en la tesis toman como entrada una ontología Datalog[±] (la cual es potencialmente inconsistente e incoherente) y retornan otra ontología Datalog[±] donde cada conflicto presente en la original fue subsanado. Siguiendo la tradición en la literatura, un requisito adicional para los procesos de consolidación presentados es que los mismos alteren las ontologías tan poco como sea posible. El enfoque utilizado se basa en *Kernel Contraction* [9, 10], el cual se centra en la resolución de conflictos mínimos conocidos como kernels. De esta forma, el proceso de resolución de conflictos es definido como la identificación de tales conflictos mínimos y la posterior remoción de alguna fórmula de los mismos para de tal forma resolver cada uno de ellos. La ontología resultante del proceso de consolidación es aquella formada por todas las fórmulas de la ontología original que no hayan sido removidas por los operadores.

2. Contribuciones y resultados de la tesis

Las principales contribuciones incluidas en la tesis versan en diferentes direcciones.

2.1. Formalización del concepto de incoherencia para Datalog[±].

Incluso cuando es un concepto muy importante para la representación de conocimiento en entornos ontológicos, la noción de incoherencia en ontologías Datalog[±] no había sido hasta el momento formalizada. Por esto, en el presente trabajo proveemos una definición formal de la misma, adaptando para tal fin una noción similar utilizada para las Lógicas Descriptivas. La formalización esta basada en la identificación de conjuntos de reglas utilizadas para la inferencia de nuevo conocimiento tales que las reglas en el conjunto no pueden ser aplicadas en su totalidad sin caer en una violación de las restricciones que se han impuesto al conocimiento. Además, ahondamos en la relación de la incoherencia con la inconsistencia, enfocándonos especialmente en como tal relación afecta la calidad de los procesos de consolidación de ontologías Datalog[±], particularmente la pérdida de información, al punto de ser necesario definir operadores que tratan ambos problemas pero de forma separada.

2.2. Establecimiento de las propiedades esperadas para procesos de consolidación de ontologías Datalog[±].

Brindamos un conjunto de propiedades que se espera sean satisfechas por operadores de consolidación de ontologías Datalog[±], adaptando las mismas a los casos de operadores trabajando a nivel local o global según el caso. Tales propiedades son presentadas a través de postulados, considerando algunas intuiciones en enfoques tradicionales para la Revisión de Creencias los cuales son adaptados al entorno ontológico de Datalog[±] (pero que pueden ser adaptados para que se adecúen a otros lenguajes ontológicos). Tales postulados proveen una caracterización formal de los operadores de consolidación sin adentrarse en *cómo* el proceso de consolidación es efectivamente realizado, proveyendo de esta forma un marco de comparación formal de operadores de consolidación de ontologías Datalog[±], tanto los definidos en la presente tesis como cualquier otro que trabaje en tales entornos.

2.3. Definición de construcciones para operadores de consolidación locales y globales.

En la tesis introducimos construcciones nóveles completas tanto para operadores de consolidación que tratan kernels por separado como para aquellos que explotan la relación de los mismos en un enfoque global. Tales construcciones toman una ontología Datalog[±] (posiblemente) inconsistente e incoherente y devuelve como resultado un subconjunto consistente y coherente de la misma. Una característica destacable de los operadores es que los mismos involucran enfoques desdoblados, donde primero el operador considera conflictos de incoherencia y resuelve los problemas de consistencia en un paso subsecuente, lo cual ayuda a mejorar el resultado final obtenido.

3. Incoherencia en ontologías Datalog[±]

Como explicamos, en el trabajo de la tesis nos hemos enfocado en la resolución de conflictos en ontologías Datalog[±]. La familia Datalog[±] de variantes de Datalog es un lenguaje que ha surgido en los últimos tiempos, y se ha vuelto popular debido a combinar una adecuada expresividad con una escalabilidad capaz de manejar los grandes volúmenes de datos presentes en las aplicaciones hoy en día. Por ejemplo, tales lenguajes extienden el Datalog básico con la posibilidad

de utilizar cuantificaciones existenciales en las cabezas de las reglas, entre otras características importantes, manteniendo a su vez una tratabilidad adecuada para entornos de alto volumen de información mediante restricciones a la sintaxis de reglas. Datalog[±] permite un estilo modular de representación de conocimiento basado en reglas, lo que lo hace útil en diversos entornos prácticos, como ser resolución de consultas en ontologías (ontology querying), extracción de datos de la Web (Web data extraction) o intercambio de datos (data exchange) [14].

La representación de conocimiento en ontologías Datalog[±] se lleva a cabo mediante el uso de (a) una Base de Datos: un conjunto de átomos que representan hechos acerca del mundo, v.g., alumno(pedro) (b) Tuple-generating Dependencies - TGDs: reglas que nos permiten obtener nuevos átomos mediante la activación de las mismas como ser: alumno(X) → persona(X), (c) Equality-generating Dependencies: reglas que restringen la generación de átomos, por ejemplo: doctor(D, P) ∧ doctor(D', P) → D = D'; y (d) Negative Constraints - NCs: reglas que expresan relaciones que no pueden existir entre átomos, v.g.,: alto(X) ∧ bajo(X) → ⊥.

Un fenómeno ampliamente estudiado en IA (incluso en Datalog[±]) es el de inconsistencia. Intuitivamente, la inconsistencia en Datalog[±] surge cuando una o más NCs o EGDs en una ontología son violadas, ya sea a través de la presencia de átomos en conflicto en el componente D o la generación de estos átomos a través de la aplicación de TGDs. Es decir, toda ontología inconsistente es tal que no tiene ningún modelo, lo que significa que no hay forma de satisfacer todas las restricciones de la misma al mismo tiempo (recordemos que un modelo es tal que toda fórmula en Σ es satisfecha).

Muchos trabajos enfocados en el manejo de inconsistencia utilizan restricciones para enfocar la resolución de inconsistencias. En general, se asume que tales restricciones no tienen problemas y por lo tanto son satisfacibles. Sin embargo, tales suposiciones no siempre son aplicables. En el trabajo desarrollado para la tesis elegimos considerar que tanto los datos como las restricciones pueden cambiar a medida que una ontología crece, y por lo tanto no sólo consideramos problemas relacionados a los datos sino también a las restricciones impuestas sobre los mismos. En este escenario más general un nuevo fenómeno aparece: incoherencia. Nuestra concepción de (in)coherencia está basada en la insatisfacibilidad de un conjunto de TGDs. Intuitivamente, consideramos un conjunto de TGDs como satisfacible cuando existe un conjunto de átomos que desencadena la aplicación de las TGDs y no produce la violación de ninguna restricción en $\Sigma_{NC} \cup \Sigma_E$, esto es, las TGDs pueden ser satisfechas junto con las NCs y las EGDs en KB .

Ejemplo 1 (Conjuntos insatisfacible de TGDs) Como un ejemplo de insatisfacibilidad (y por lo tanto incoherencia) considere el siguiente conjunto de restricciones:

$$\Sigma_{NC}^1 : \{\tau_1 : garganta_lastimada(X) \wedge puede_cantar(X) \rightarrow \perp\}$$

$$\Sigma_E^1 : \{\nu_1 : representa(X, Y) \wedge representa(X, Z) \rightarrow Y = Z\}$$

$$\Sigma_T^1 : \{\sigma_1 : cantante_rock(X) \rightarrow canta_fuerte(X), \\ \sigma_2 : canta_fuerte(X) \rightarrow garganta_lastimada(X), \\ \sigma_3 : cantante_rock(X) \rightarrow puede_cantar(X)\}$$

El conjunto Σ_T^1 es un conjunto insatisfacible de dependencias, ya que la aplicación de las TGDs $\{\sigma_1, \sigma_2, \sigma_3\}$ en cualquier conjunto relevante de átomos (cualquier conjunto que dispare todas las TGDs) causará la violación de τ_1 , v.g., si consideramos el átomo relevante cantante_rock(axl) tenemos que Axl puede cantar y tiene la garganta lastimada. Nótese que cualquier conjunto relevante de átomos (esto es, cualquier cantante que consideremos) causará la violación de τ_1 .

Bajo el uso de este nuevo concepto de satisfacibilidad de un conjunto de TGDs desarrollado para la tesis tenemos que una ontología es coherente en el caso de que el conjunto de TGDs en la misma es satisfacible con respecto a las EGDs y NCs en ella, e incoherente en el caso contrario. Como veremos más adelante, la cercana relación entre inconsistencia e incoherencia hace que

los operadores de consolidación de ontologías pueden beneficiarse de atacar no sólo problemas de inconsistencia sino también eliminando las incoherencias en el conocimiento expresado en las ontologías Datalog[±].

4. Consolidación de ontologías Datalog[±]

Como uno de los aportes principales de la tesis hemos presentado un enfoque que posibilita la consolidación de ontologías Datalog[±]. Tal proceso está basado en las ideas de Hansson [9, 10] de atacar conflictos mínimos conocidos como kernels, removiendo de los mismos fórmulas para resolver los conflictos. Además de atender conflictos de inconsistencia, los operadores definidos también se enfocan en el otro tipo de conflictos que pueden surgir en entornos ontológicos: la incoherencia. De esta forma puede verse a los operadores como unos que trabajan en dos fases: los mismos comienzan por resolver los problemas de incoherencia que aparezcan en la ontología mediante la remoción de TGDs, para luego dar paso a la restauración de la consistencia a través de la eliminación de átomos del componente D . En ambos casos, las selecciones se hacen a través de funciones de incisión que deciden qué remover en base a una relación general \prec , mirando para ello localmente en cada kernel en la ontología. Esto es, los operadores eligen de cada kernel aquella fórmula que es señalada por \prec como la menos deseable dentro del kernel.

Para caracterizar apropiadamente el proceso de consolidación de ontologías hemos introducido un conjunto de postulados, adaptando intuiciones introducidas para otros formalismos de representación de conocimiento. A tales propiedades las hemos aumentado con los postulados de Optimidad Local para Restricciones y Optimidad Local para Datos, los cuales son utilizados para enfocarnos en aquellas consolidaciones que minimizan la pérdida de información con respecto a \prec , donde la misma es considerada localmente, *esto es*, para cada conjunto minimalmente inconsistente/incoherente. El conjunto básico de Postulados presentados versa entonces en diferentes aspectos de la consolidación de ontologías, desde requerir que nada sea agregado a la ontología original hasta que aquello que no involucra conflictos sea retenido. Este conjunto básico de postulados implica a su vez ciertas otras propiedades. En particular, dos propiedades que se derivan del conjunto de postulados básicos y que son muy importantes a efectos del objetivo de los operadores de consolidación son los de **Coherencia** y **Consistencia**. Es importante remarcar que, como es tradicional en la teoría de cambios, los operadores no se encuentran atados a las construcciones particulares que se presentan en la presente tesis, sino que son un aporte general a cualquier operador de consolidación de ontologías Datalog[±] (y que pueden ser incluso generalizados a otros lenguajes ontológicos con relativa sencillez). Es decir, cualquier operador de consolidación de ontologías Datalog[±] puede ser analizado a la luz de los postulados presentados, y no solamente aquellos introducidos en este trabajo. Por lo tanto, el conjunto de postulados introducidos corresponde a un marco formal general del comportamiento de operadores de consolidación.

Luego de presentar los postulados que moldean el comportamiento de los operadores de consolidación de ontologías Datalog[±] nos hemos dedicado a la definición de una clase particular de tales operadores, presentando para ello una construcción completa que puede generarlos. Para ello primeramente procedimos a definir que corresponden con kernels de dependencias (conjuntos insatisfacibles de TGDs mínimos bajo inclusión conjuntista) y kernels de datos (conjuntos mínimos de átomos que hacen a la ontología inconsistente). En particular, para los kernels de dependencias hemos demostrado que los mismos son independientes de la instancia particular del componente D en la ontologías, es decir que los mismos no dependen de los átomos. Esto es particularmente interesante porque nos da la pauta de que el conjunto de kernels de dependencias de una ontología puede ser obtenido sin considerar el componente D de la misma, lo que es de utilidad a la hora de separar la resolución de incoherencias de la resolución de inconsistencias.

Una vez que hemos identificado los kernels sobre los que aplicaremos las operaciones de contracción en pos de resolver incoherencias e inconsistencias, debemos establecer como tales problemas son resueltos. Para ello utilizamos funciones de incisión, que seleccionan qué TGDs

y qué átomos serán removidos de los kernels de dependencias y kernels de datos, respectivamente. Es posible definir funciones de incisión generales, que consideran al mismo tiempo tanto incoherencia como inconsistencia. Sin embargo, tal enfoque acarrea un problema: para hacer esto deberíamos calcular los kernels de dependencias y de datos y resolver los conflictos al mismo tiempo. En la tesis mostramos que en la presencia de incoherencia esto puede no ser la mejor solución, al considerarlo a la luz de la pérdida mínima de información, ya que para el caso de átomos relevantes a conjuntos insatisfacibles de TGDs los mismos deben ser obligatoriamente eliminados por una función de incisión general, incluso cuando al mismo tiempo la función de incisión resuelve el problema de incoherencia que hacía que el átomo en cuestión sea un kernel (resolviendo indirectamente la inconsistencia).

Para resolver tal situación es que hemos introducido una separación entre las funciones de incisión que trabajan sobre el conjunto de TGDs de una ontología de aquellas que se encargan de la remoción de átomos de la misma. En base a tal separación definimos a los operadores de consolidación de ontologías $Datalog^\pm$ como aquellos que utilizan una función de incisión en restricciones sobre el componente Σ y una función de incisión en datos sobre D . Se puede considerar que tales operadores trabajan en “fases” separadas: primero se resuelven todos los problemas de incoherencia, y luego la resolución de inconsistencias se hace sobre la ontología intermedia obtenida mediante la remoción de TGDs. Es decir, por un lado, las TGDs son removidas de Σ (solamente TGDs, ya que los kernels de dependencias no contienen EGDs ni NCs). Por el otro lado, la función de incisión en datos usa una ontología intermedia KB^* en lugar de KB . De esta forma logramos que sólo sean removidos átomos de D que estén en conflicto con el conjunto de TGDs resultante de resolver las incoherencias, ya que los kernels de datos son calculados en base a las restricciones obtenidas luego de aplicar sobre Σ la función de incisión en restricciones.

Por último, hemos establecido la relación entre los postulados y la construcción presentada a través de un teorema de representación, el cual establece que la relación biunívoca entre el conjunto de postulados OCP1-OCP7 y la construcción en base a funciones de incisión. La importancia de tal resultado es que determina que cualquier construcción que satisfaga tales postulados se corresponderá necesariamente con los operadores presentados en este capítulo, y por lo tanto los resultados de esos operadores pueden ser obtenidos por nuestra construcción.

5. Refinamiento del proceso de consolidación de ontologías $Datalog^\pm$

Además de los operadores locales mencionados anteriormente, en la tesis hemos introducido una clase alternativa de operadores de consolidación de ontologías que modifican el comportamiento los mencionados. La diferencia entre los enfoques se da en que, mientras que los operadores presentados previamente atacan localmente a cada conflicto mínimo, en el enfoque alternativo se utiliza una mirada más global, considerando la relación (cuando la misma existe) entre los distintos kernels. La consideración de tal relación se hace mediante el agrupamiento de kernels relacionados en una superestructura denominada cluster, la cual es obtenida explotando una relación de overlapping entre conflictos.

Para ello primeramente hemos analizado el comportamiento de los operadores locales, enfocando el análisis en la pérdida innecesaria de información que los mismos pueden sufrir bajo ciertas condiciones, a saber la existencia de kernels relacionados y estrategias de resolución de conflictos donde la elección óptima local en cada kernel hace que debamos de remover de un kernel una fórmula (la indicada por la relación de preferencia) incluso cuando otra fórmula del mismo kernel ya ha sido seleccionada para remoción como estrategia de solución del conflicto expresado por un kernel diferente. Para poder ver lo expresado considere el siguiente ejemplo, que ilustra tanto el comportamiento de un operador local como el problema que puede traer aparejado.

Ejemplo 2 Supongamos que tenemos $\perp\!\!\!\perp_{(D,KB^*)} = \{\{a_1, a_3\}, \{a_2, a_3\}\}$, y \prec donde $a_1 \prec a_3 \prec a_2$, donde $\perp\!\!\!\perp_{(D,KB^*)}$ denota el conjunto de kernels de datos en la ontología. Entonces, tenemos que la función de incisión ϱ es tal que

$$\begin{aligned}\varrho(\{a_2, a_3\}) &= \{a_3\} \\ \varrho(\{a_1, a_3\}) &= \{a_1\}\end{aligned}$$

La remoción de a_1 es innecesaria, ya que el conflicto es resuelto indirectamente mediante la remoción de a_3 del primer kernel. Sin embargo, un operador local lo removería, ya que no considera la interacción entre los kernels. Dados estos dos kernels las incisiones minimales son o bien remover el átomo a_3 o bien remover $\{a_1, a_2\}$.

Esta situación pone en evidencia que el conjunto de postulados introducidos previamente no es suficiente para paliar completamente el problema de pérdida mínima de información, por lo que hemos introducido una propiedad adicional. La propiedad que se espera sea satisfecha por un operador optimal (en el sentido de mínimo cambio) es la de **Mínima Pérdida de Información**. La intuición detrás del postulado es que un operador de consolidación que lo satisface nos dará como resultado una ontología tal que si agregamos cualquiera de las fórmulas eliminadas entonces la KB resultante sería incoherente o inconsistente; es decir, el operador realizó una cantidad de cambios *mínima* en la resolución de conflictos.

Como se ve en Ejemplo 2, la propiedad de mínima pérdida no es satisfecha por los operadores locales, debido al uso de kernels de manera aislada, sin tener en cuenta aquellas ocasiones donde los conflictos minimales están en una relación de “colisión”. Es por esto que en la tesis introducimos una construcción alternativa que se basa en la detección de la relación entre kernels, para de esta forma encontrar la solución que implique la menor pérdida posible. Para lograr esto, en lugar de realizar incisiones sobre kernels el nuevo operador realiza las mismas sobre *clusters*. La estructura de los clusters hace posible identificar conflictos relacionados a través de una relación de *solapamiento* (overlapping) entre kernels.

Sin embargo, el uso de clusters trae aparejado un problema: al remover alguna fórmula de un kernel el conjunto resultante siempre es coherente/consistente ya que los kernels son conjuntos mínimos, pero en el caso de los clusters no podemos asegurar que la remoción de una fórmula nos de como resultado un conjunto con todos los conflictos agrupados resueltos [14]. Para paliar este problema hemos presentado una nueva clase de funciones de incisión que aseguran que una vez que el conjunto de fórmulas seleccionado por la función es removido todos los conflictos mínimos incluidos en el cluster quedan resueltos. En base a tales funciones de incisión hemos definido entonces a los operadores de consolidación de ontologías Datalog^\pm basados en Cluster Contraction los cuales, al igual que los operadores locales, utilizan remoción de átomos y TGDs como estrategia de resolución de incoherencias e inconsistencias.

Una vez presentada la construcción propuesta para los operadores de consolidación de ontologías basados en Cluster Contraction hemos ahondado en la relación existente entre tal construcción y las propiedades presentadas, a través de un teorema de representación. Tal teorema muestra que estos operadores efectivamente satisfacen el postulado de mínima pérdida de información, y por lo tanto representación soluciones óptimas desde tal punto de vista. Es decir, que los operadores basados en Cluster Contraction son tales que remueven lo mínimo que es requerido para solucionar los conflictos, y por lo tanto si fuéramos a dejar siquiera una fórmula de aquellas removidas por el operador entonces la ontología seguiría siendo incoherente/inconsistente.

Un hallazgo interesante presentado en la tesis es que existen casos donde los enfoques coinciden. Un claro caso (trivial) de ésta situación se da al tratar con KB s coherentes y consistentes. Sin embargo, no es el único caso donde tal cosa puede suceder. Otro caso que llevaría a lo mismo es aquél donde los conjuntos de kernels y de clusters coinciden, *esto es*, no hay overlapping entre conflictos mínimos; nuevamente, para tales casos los enfoques coincidirían. Esto, si bien puede parecer trivial y no reportar mucho interés en una primera mirada, se vuelve importante a la hora

de analizar a los operadores desde un punto de vista computacional. Claramente, el manejo de kernels es más sencillo que el manejo de clusters; por ejemplo, como remover una fórmula alcanza para solucionar conflictos mínimos entonces podemos comparar conjuntos unitarios de fórmulas, sin necesidad de mirar en todos los subconjuntos como sucede en Cluster Contraction. Por lo tanto, si tenemos que el conjunto de clusters coincide con el de kernels (lo que sólo involucraría una comparación de conjuntos adicional) podríamos directamente tomar el enfoque local, consiguiendo una mejor performance computacional. Es más, esto puede ser llevado un paso más allá, definiendo operadores mixtos: podríamos utilizar un enfoque local para el caso donde *un* cluster coincide con algún kernel, y utilizar un enfoque global sobre aquellos que involucran dos o más conflictos mínimos.

6. Conclusiones

El trabajo colaborativo y el intercambio de información se han transformado en los últimos tiempos en aspectos cruciales de prácticamente todo sistema; por lo tanto, es de vital importancia disponer de métodos automáticos y adecuados para el manejo de conflictos: en entornos colaborativos, a medida que el conocimiento evoluciona inconsistencias e incoherencias suelen aparecer de manera natural. Es más, en entornos colaborativos tal conocimiento es frecuentemente representado a través de ontologías que suelen ser mantenidas en forma conjunta por muchas entidades, *esto es*, son compartidas por entidades que no sólo las utilizan sino que también las modifican.

Una forma de lidiar con los conflictos que pueden aparecer en tales entornos de aplicación es tratar de modificar la información contenida en la ontología para de esta forma recuperar la coherencia y la consistencia del conocimiento expresado por las mismas. En este trabajo hemos presentado una forma de conseguir la consolidación de ontologías Datalog^{\pm} , *esto es*, la resolución de todos los conflictos presentes en las mismas, tanto de inconsistencia como de incoherencia.

En particular, el fenómeno de incoherencia es uno que suele ser dejado de lado, a pesar de tener gran influencia en la definición de procesos de manejo de inconsistencia. Para paliar esto hemos presentado una caracterización del fenómeno de incoherencia en Datalog^{\pm} , tomando como punto de partida esfuerzos similares en DLs, donde se relaciona a tal fenómeno con la aparición de conceptos insatisfacibles [6, 5, 3, 13, 17, 16].

Además de introducir los conceptos de satisfacibilidad y coherencia para Datalog^{\pm} , en el presente trabajo hemos ahondado en la relación entre inconsistencia e incoherencia. En particular, el trabajo muestra que tal relación toma la forma de causa-efecto: cuando tenemos una ontología incoherente la misma es inevitablemente inconsistente cuando el componente D hace que un conjunto insatisfacible sea activado. Esta relación es de vital importancia para la definición de procesos de manejo de inconsistencia: es fácil demostrar que en presencia de una ontología incoherente puede ser una mejor solución aplicar un enfoque de eliminación de incoherencia como paso previo al manejo de inconsistencia, lo que permitirá aumentar la cantidad de respuestas posibles para la ontología previamente incoherente.

De la mano de diferenciar y atacar de forma correspondiente cada tipo de conflicto es que en la presente tesis hemos propuesto la utilización de operadores de consolidación de ontologías Datalog^{\pm} que tengan en cuenta ambas clases de conflictos, *esto es*, que ataquen tanto incoherencia como inconsistencia. Esto último explicado es central dentro de la construcción de operadores de consolidación propuestos, ya que la misma está basada en la separación en dos tipos de funciones de incisión de modo que se pueda atacar primeramente los problemas de coherencia utilizando una función de incisión en restricciones ρ , y luego resolver los problemas de consistencia basándonos en el resultado de remover de la ontología original aquellas TGDs seleccionadas por ρ . Esto es; el resultado de una operación de consolidación $\Upsilon_{\rho, \varrho}(KB)$ es la ontología obtenida removiendo TGDs (seleccionadas por una función de incisión en restricciones ρ del conjunto de kernels de dependencias $\perp\!\!\!\perp_{(\Sigma, KB)}$) y átomos (seleccionados por una función de incisión en datos ϱ de $\perp\!\!\!\perp_{(D, KB^*)}$, que es el conjunto de kernels de datos para la ontología resultante de resolver las

incoherencias como primer paso) de la ontología original *KB*.

La primer construcción propuesta ataca directamente kernels de manera local, lo que acarrea un problema al examinarla a la luz de la mínima pérdida de información; en ocasiones, cuando dos o más kernels están relacionados su tratamiento local no es óptimo. Para paliar tal problema es que hemos definido una estructura adicional denominada cluster, que tiene la función de agrupar kernels relacionados de modo que podamos resolver tales conflictos utilizando un enfoque global del problema. Sin embargo, la aplicación de funciones de incisión definidas para trabajar en kernels directamente a clusters no fue posible, por lo que se presentan nuevas funciones de incisión adaptadas a su uso en clusters, y se define una nueva clase de operadores de consolidación que aseguran que siempre la ontología resultante de la consolidación es la más cercana a la ontología original (en términos de pérdida de información), aunque como contrapartida el esfuerzo computacional puede llegar a ser mayor.

Paralelamente a la presentación de las construcciones antes mencionadas, en la presente tesis hemos presentado una serie de propiedades que operadores de consolidación de ontologías *Datalog*[±] deberían respetar para exhibir un comportamiento que preserve los requisitos de resolver los conflictos que aparecen en las ontologías, tratando a su vez de minimizar las pérdida de información. Como es característico en Revisión de Creencias, estas propiedades son presentadas como postulados, donde los mismos son independientes de la construcción propuesta; esto es, los postulados pueden ser utilizados para medir el desempeño de cualquier operador de consolidación de ontologías *Datalog*[±], y no solamente aquellos propuestos en este trabajo.

Respecto de la relación entre las construcciones presentadas (local/kernel y global/cluster) y las propiedades esperadas que expresan los postulados, hemos establecido claramente la relación entre ambos aportes mediante distintos teoremas de representación. El aporte de tales teoremas es doble: (a) muestran claramente como las construcciones propuestas satisfacen las propiedades esperadas para cada tipo de operador, y (b) muestran que cada operador de consolidación que sea definido de forma tal que satisfaga las propiedades en los teoremas se corresponderá necesariamente con alguna de las construcciones propuestas en este trabajo, es decir, que dada la misma ontología original obtendrá como resultado la misma ontología consolidada que se obtiene utilizando las construcciones presentadas .

7. Líneas de investigación futuras

Obtención de vistas unificadas de varias bases de datos relacionales.: Una aplicación posible de los operadores de consolidación de ontologías *Datalog*[±] es su utilización para la consecución de una vista única de diferentes fuentes de información expresables por *Datalog*[±] (como por ejemplo Bases de Datos Relacionales), mediante un proceso de transformación de las bases de datos relacionales a ontologías *Datalog*[±] primeramente, para luego consolidar (la unión de) las mismas y obtener la vista.

Uso de la consolidación como soporte a procesos de evolución de ontologías: La evolución de ontologías (ontology evolution) trata de modificar las ontologías para reflejar cambios en el dominio que las mismas representan de forma tal que no haya conflictos en la ontología resultante. Para conseguir esto debemos definir procesos de revisión en *Datalog*[±] mediante el enfoque de consolidación presentado como la base del desarrollo de operadores de revisión de ontologías *Datalog*[±]. Para esto, tales operadores comienzan agregando la fórmula en consideración (ya sea un átomo o una TGD) para luego asegurarse de que lo resultante de tal expansión no sea inconsistente/incoherente mediante la remoción de átomos o TGDs. En base a (clases particulares de) tales operadores podemos definir un proceso de evolución de ontologías como aquel que revisa a la ontología actual con el nuevo átomo a agregar, logrando de esta forma incluir esta nueva información sin que aparezcan por ellos conflictos en la ontología resultante.

Manejo de incoherencias a través de debilitación de reglas: Defeasible *Datalog*[±]: Una propuesta alternativa para el manejo de incoherencias que pensamos explorar es la basada en de-

bilitamiento de reglas, en lugar de remoción de las mismas. Para lograr esto pensamos explotar unos desarrollos recientes que han surgido dentro del grupo de investigación que anidó el trabajo presentado en esta disertación, pero que se centran en la modificación de la relación de consecuencia para el manejo de ontologías inconsistentes: Datalog[±] rebatible [15].

Implementación de operadores: Finalmente, en la actualidad estamos trabajando en la implementación de nuestros operadores; planeamos para esto estudiar diferentes técnicas que puedan ser utilizadas para obtener una implementación eficiente de los operadores, posiblemente enfocados en fragmentos específicos de Datalog[±].

Referencias

- [1] L. Amgoud and S. Kaci. An argumentation framework for merging conflicting knowledge bases: The prioritized case. In *ECSQUARU 2005*, pages 527–538, 2005.
- [2] D.A. Bell, G. Qi, and W. Liu. Approaches to inconsistency handling in description-logic based ontologies. In *OTM Workshops*, pages 1303–1311, 2007.
- [3] D. Beneventano and S. Bergamaschi. Incoherence and subsumption for recursive views and queries in object-oriented data models. *DKE*, 21(3):217–252, 1997.
- [4] E. Black, A. Hunter, and J. Z. Pan. An argument-based approach to using multiple ontologies. In *SUM 2009*, pages 68–79, 2009.
- [5] A. Borgida. Description logics in data management. *TKDE*, 7(5):671–682, 1995.
- [6] G. Flouris, Z. Huang, J. Z. Pan, D. Plexousakis, and H. Wache. Inconsistencies, negations and changes in ontologies. In *AAAI 2006*, pages 1295–1300, 2006.
- [7] S. A. Gómez, C. I. Chesñevar, and G. R. Simari. Reasoning with inconsistent ontologies through argumentation. *App Artif. Intell.*, 24(1&2):102–148, 2010.
- [8] P. Haase, F. van Harmelen, Z. Huang, H. Stuckenschmidt, and Y. Sure. A framework for handling inconsistency in changing ontologies. In *ISWC 2005*, pages 353–367, 2005.
- [9] S. O. Hansson. Kernel contraction. *J. of Symb. Log.*, 59(3):845–859, 1994.
- [10] S. O. Hansson. *A Textbook of Belief Dynamics: Theory Change and Database Updating*. Kluwer Academic Publishers, Norwell, MA, USA, 1999.
- [11] Sven Ove Hansson. *Belief Base Dynamics*. PhD thesis, Uppsala University, Department of Philosophy, Uppsala, Sweden, 1991.
- [12] Z. Huang, F. van Harmelen, and A. ten Teije. Reasoning with inconsistent ontologies. In *IJCAI 2005*, pages 454–459, 2005.
- [13] A. Kalyanpur, B. Parsia, E. Sirin, and J. A. Hendler. Debugging unsatisfiable classes in owl ontologies. *Web Semantics: Science, Services and Agents on the World Wide Web*, 3(4):268–293, 2005.
- [14] T. Lukasiewicz, M. V. Martínez, and G. I. Simari. Inconsistency handling in datalog+/- ontologies. In *ECAI 2012*, pages 558–563, 2012.
- [15] M. V. Martínez, C. A. D. Deagustini, M. A. Falappa, and G. R. Simari. Inconsistency-tolerant reasoning in datalog[±] ontologies via an argumentative semantics. In *IBERAMIA 2014*, pages 15–27, 2014.
- [16] G. Qi and A. Hunter. Measuring incoherence in description logic-based ontologies. In *ISWC/ASWC 2007*, pages 381–394, 2007.
- [17] S. Schlobach, Z. Huang, R. Cornet, and F. van Harmelen. Debugging incoherent terminologies. *J. of Autom. Reason.*, 39(3):317–349, 2007.