



REVECOM

Revista Venezolana de Computación

Sociedad Venezolana
de Computación

ISSN: 2244-7040

Vol. 1, No. 1
Julio 2014



REVECOM

Revista Venezolana de Computación

Sociedad Venezolana
de Computación

Editores:
Eric Gamess, Wilmer Pereira, Yudith Cardinale

ISSN: 2244-7040

Vol. 1, No. 1
Julio 2014

Editorial

El Nacimiento de la Revista Venezolana de Computación

ReVeCom (Revista Venezolana de Computación) es la primera revista venezolana arbitrada orientada a la publicación de resultados de investigación en el campo de la computación. ReVeCom fue creada por la SVC (Sociedad Venezolana de Computación) y tiene entre sus objetivos hacer conocer los trabajos de alta calidad investigativa que se realizan a nivel nacional, latinoamericano e internacional. Adicionalmente, cada año, se editará un número especial de ReVeCom con una selección de los mejores artículos presentados en CoNCISa (Conferencia Nacional de Computación, Informática y Sistemas).

ReVeCom es un instrumento digital de difusión científica, sin fines de lucro, para la expansión del conocimiento en el área de la computación. ReVeCom utiliza el medio digital para hacer más amplia su difusión y permitir un fácil acceso a los investigadores, tanto a nivel nacional como internacional. Para su administración, la revista se basa en el estándar OJS (Open Journal System), ya que es utilizado en la mayoría de los repositorios de "Saber" de las universidades en Venezuela. La revista permite la publicación de artículos con aporte original en castellano o inglés.

En la actualidad, sólo existen en el país algunas revistas científicas impresas por las bibliotecas y editoriales de las más importantes universidades nacionales públicas y privadas. Estas revistas ya existentes cubren un área grande de las ciencias y/o ingeniería, no siendo específicas al área de computación. En contra parte, la idea es que ReVeCom aporte visibilidad al utilizar los medios digitales y ser promovida por la SVC, rápida difusión y edición, además de tener la virtud de centrarse en un público especializado.

La edición de este primer número de ReVeCom está dedicada a los mejores trabajos presentados en CoNCISa 2013, celebrada en el Club Puerto Azul, Naiguatá, Venezuela. Esta edición consolida un esfuerzo grande que se ha venido haciendo en el seno de la SVC, para promover la investigación en el campo de la computación a nivel nacional, e impulsar una nueva generación académica y profesional en nuestra área de saber para el desarrollo del país.

Finalmente, al nombre del comité editorial de ReVeCom, quiero agradecer a los autores que contribuyeron a la edición de este primer número de ReVeCom con la publicación de sus resultados de investigación, y hacer una invitación amplia a la comunidad de computación a someter sus propios trabajos para los números de ReVeCom por venir.

Dr. Eric Gamess
Director de la Revista Venezolana de Computación
Universidad Central de Venezuela

Sociedad Venezolana de Computación

La SVC (Sociedad Venezolana de Computación) tiene como objetivo principal auspiciar actividades académicas, científicas y profesionales, en pro del desarrollo de la computación y carreras afines, en Venezuela. Su primera junta directiva está formada por profesores de importantes universidades nacionales cuya motivación inicial es estructurar, bajo un marco legal, actividades académico-científicas que perduren en el tiempo.

Anecdóticamente comenzó a gestarse en el exterior, durante la reunión de la junta directiva del CLEI (Centro Latinoamericano de Estudios en Informática) en el 2011. Esa reunión tuvo lugar en Quito, Ecuador, en el marco de la Conferencia Latinoamericana de Informática. Allí el representante por Venezuela ante CLEI, en aquel momento el Profesor Wilmer Pereira, habló con colegas latinoamericanos, específicamente con los fundadores de las Sociedades de Computación Peruana y Colombiana, sobre su experiencia. Gracias a ese intercambio de ideas, se estableció una primera versión de los estatutos que fueron revisados en mayo de 2012, durante el SCTC 2012 (Simposio Científico y Tecnológico en Computación 2012) organizado por la Escuela de Computación de la Universidad Central de Venezuela. En noviembre de ese mismo año, durante la realización de la LXII Convención Anual de la AsoVAC, organizada por la Universidad Metropolitana, se acordó registrar la SVC como asociación sin fines de lucro y se conformaron las primeras comisiones de trabajo.

Una de esas comisiones asumió la organización de la CoNCISa 2013 (Conferencia Nacional de Informática, Computación y Sistemas 2013). Aprovechando la realización de la XXXIX Conferencia Latinoamericana de Informática que se iba a celebrar en Venezuela, se organizó en paralelo CoNCISa 2013 con la participación de varias universidades nacionales de Venezuela (UCV, UCAB, USB, UMC y UBV), para beneficiarse de la presencia de investigadores y tutorialistas de toda latinoamérica.

Otra iniciativa indispensable en Venezuela para el desarrollo del sector académico en computación, es la Revista Venezolana de Computación (ReVeCom). ReVeCom tiene en sus objetivos hacer conocer los trabajos de alta calidad investigativa que se realizan a nivel nacional, latinoamericano e internacional. Adicionalmente, una selección de los mejores artículos presentados en CoNCISa son invitados a ser publicados en ReVeCom.

La Sociedad Venezolana de Computación tiene además otros proyectos como:

1. Estudio de la propuesta ACM/IEEE para carreras afines a computación, informática y sistemas. El objetivo no es adoptar literalmente alguno de los cinco planes de estudio propuestos, sino poder situar los pregrados de las universidades venezolanas en esta clasificación. Así se facilitarán los procesos de equivalencias y homologación con universidades nacionales e internacionales. Es claro que nuestra realidad exige a las universidades venezolanas un perfil estudiantil particular. Sin embargo, no podemos estar desligados del resto del mundo por lo que es imprescindible poder compararnos con los perfiles ACM/IEEE.
2. Apoyo a eventos estudiantiles como maratones de programación, jornadas estudiantiles, competencia de robótica. Las actividades estudiantiles aumentan en cantidad y variedad y son una fuente de motivación importante para proyectos académicos y científicos. Es por ello que la SVC tiene la firme intención de apurar estas iniciativas.

3. Conformación de un postgrado nacional, aprovechando las herramientas de e-learning y de educación virtual que ya tienen muchas universidades nacionales. Este es el proyecto más reciente y apenas se están estructurando los grupos de trabajo.

En fin, la SVC quiere mantenerse en el tiempo con proyectos indispensables para el desarrollo de las ciencias de la computación en Venezuela, sirviendo de marco legal y, de ser necesario, regulatorio a nivel nacional. Como en toda sociedad, existe un procedimiento de membresía que ofrece ventajas a quienes estén adscritos para participar en las distintas actividades organizadas o patrocinadas por la SVC. Invitamos al lector no sólo a hacerse miembro, sino a participar en los distintos proyectos de la SVC. Para más información pueden consultar <http://www.svc.net.ve> donde encontrarán detalles de nuestras actividades.

Dr. Wilmer Pereira
Presidente de la Sociedad Venezolana de Computación
Universidad Católica Andrés Bello

Directorio de la Sociedad Venezolana de Computación

Presidente:

Dr. Wilmer Pereira (Universidad Católica Andrés Bello)

Secretario:

Dr. Carlos Acosta (Universidad Central de Venezuela)

Tesorero:

Dr. Carlos Figueira (Universidad Simón Bolívar)

Vicepresidente:

Profesor Robinson Rivas (Universidad Central de Venezuela)

Coordinador de Eventos:

Dr. Leonid Tineo (Universidad Simón Bolívar)

Coordinadora de Publicaciones:

Dra. Judith Barrios (Universidad de los Andes)

Coordinadora de Educación e Investigación:

Dra. Yudith Cardinale (Universidad Simón Bolívar)

Edición

Comité Editorial

Director:

Dr. Eric Gamess - Universidad Central de Venezuela, Venezuela
Redes de computadores, computación de alto desempeño, simulación.

Coordinador del Comité Editorial:

Dr. Wilmer Pereira - Universidad Católica Andrés Bello, Venezuela
Inteligencia artificial, robótica autónoma, aprendizaje automatizado.

Jefe de Redacción:

Dra. Yudith Cardinale - Universidad Simón Bolívar, Venezuela
Computación paralela, computación de alto desempeño, sistemas distribuidos, computación en la nube, arquitecturas paralelas, servicios web, web semántica.

Miembros del Comité Editorial

Dr. Carlos Acosta - Universidad Central de Venezuela, Venezuela
Computación paralela, computación de alto desempeño, computación reconfigurable y FPGAs, simulación paralela y distribuida, BigData.

Dr. Andrés Arcia-Moret - Universidad de los Andes, Venezuela
Simulación de redes, protocolos de transporte, redes inalámbricas.

Dr. Ernesto Coto - The University of Sheffield, Inglaterra
Computación gráfica, visualización científica, procesamiento digital de imágenes.

Dra. Francis Losavio - Universidad Central de Venezuela, Venezuela
Ingeniería del software, arquitecturas y calidad del software, producción industrial de software.

Dr. Francisco Luengo - Universidad del Zulia, Venezuela
Computación social, minería de texto.

Dr. Jonas Montilva - Universidad de los Andes, Venezuela
Ingeniería del software, sistemas de información.

Dra. Masun Nabhan - Universidad Simón Bolívar, Venezuela
Inteligencia artificial, minería en datos, aplicaciones de inteligencia artificial para educación y discapacitados.

Dra. Dinarle Ortega - Universidad de Carabobo, Venezuela
Ingeniería del software, arquitectura del software, arquitecturas empresariales, modelado de procesos de negocio.

Dr. David Padua - University of Illinois, USA
Compiladores, computación de alto desempeño.

Dr. Leonid Tineo - Universidad Simón Bolívar, Venezuela
Bases de datos, lógica difusa, lenguajes artificiales, minería de datos.

Tabla de Contenido

1.	Estudio sobre Competencias del Ingeniero en Informática: Caso Universidad Católica Andrés Bello	1-7
	Susana García, Wilmer Pereira, Rosaura Paladino, Lucia Cardoso, Lisset de Gouveia	
2.	Un Framework para la Rehabilitación Física en Miembros Superiores con Realidad Virtual	8-16
	Esmitt Ramírez, Francisco Moreno, Jordan Ojeda, Christiam Mena, Omaira Rodríguez, Jeanlight Rangel, Silvio Álvarez	
3.	Revisión Documental Sistemática de Enfoques y Técnicas para la Construcción de Arquitecturas en un Contexto de Líneas de Productos de Software	17-25
	Juan Herrera, Alfredo Matteo, Francisca Losavio	
4.	Desarrollo de Aplicaciones Usando Modelos de Procesos Flexibles	26-33
	Maria Guevara, Dinarle Ortega, Argenis Guerrero	
5.	Retos para el Procesamiento Semántico de Datos Enlazados en la Nube de los Datos Abiertos Enlazados	34-42
	Maria-Esther Vidal, Maribel Acosta, Ana Alvarado, Oriana Baldizán, Alexander Baranya, Simón Castillo, Giuseppe De Simone, Marlene Goncalves, Hancel González, Alexandra La Cruz, Gabriela Montoya, Guillermo Palma, Edna Ruckhaus	
6.	OCTOPUS: Aplicación de SQLf para Selección de Fuentes de Datos	43-49
	Marlene Goncalves, Leonid Tineo	
7.	Identificación de Secuencias de Scanning en Redes 802.11	50-57
	Antonio Araujo, Andrés Arcia-Moret	
8.	Gestión de Tráfico Best Effort a través de Retardos Aleatorios en Redes 802.16	58-64
	Andrés Arcia-Moret, Diego Uzcátegui	
	Índice de Autores	65

Estudio sobre Competencias del Ingeniero en Informática: Caso Universidad Católica Andrés Bello

Susana García¹, Wilmer Pereira^{1,2}, Rosaura Paladino¹, Lucia Cardoso¹, Lisset de Gouveia¹
sgarcia@ucab.edu.ve, wpereira@usb.ve, rpaladin@ucab.edu.ve, lgouveia@ucab.edu.ve

¹ Universidad Católica Andrés Bello, Montalbán, Caracas, Venezuela

² Departamento de Computación, Universidad Simón Bolívar, Caracas, Venezuela

Resumen: En el año 2008 la UCAB (Universidad Católica Andrés Bello) inició un proceso de renovación curricular por competencias, bajo los lineamientos de la Comisión de Curriculum Institucional, usando como base el proyecto *Tuning*. La UCAB, desde las más altas instancias, definió las competencias generales de sus egresados y, a partir de esa lista, un grupo de profesores de informática definió las competencias del ingeniero informático, como proyecto piloto, de la Facultad de Ingeniería. Inicialmente se obtuvieron 24 competencias y con ellas se aplicó una encuesta a buena parte del cuerpo profesoral, algunos empleadores y una cierta cantidad de estudiantes de último semestre y egresados. Esta población indicó, para cada competencia, el nivel importancia y que tanto la desarrollan los egresados. Tomando como insumo los resultados de esta encuesta, se aplicó un análisis factorial, y surgieron 4 dimensiones que permitieron afinar el perfil del egresado. Acto seguido se asignaron las respectivas unidades de competencia, por cada dimensión, que todos los graduados en ingeniería informática deben lograr para su ejercicio profesional. Por último, gracias a esta primera versión de las unidades de competencia del ingeniero en informática, se definieron las competencias generales del ingeniero de la facultad incluidas las carreras de telecomunicaciones, industrial y civil.

Palabras Clave: Curricula por Competencias; Proyecto Tuning; EEES; Ingeniería Informática.

Abstract: In 2008 the Andrés Bello Catholic University (UCAB) initiated a process of curricula reform under the guidelines of the Committee on Institutional Curriculum, using as a basis the Tuning project. Authorities university defined the general skills of their graduates, and a group of computer teachers defined a first version on curricula of computer engineering. Initially 24 skills were obtained and a survey was applied to the teaching staff, some employers and a certain amount of students and graduates. This population indicated, for each skill, a level importance for professionals in computer engineering. Taking the results, a factor analysis was applied, and emerged four dimensions that allowed tuning the professional profile of computer engineer of UCAB. Each skill were assigned to a dimension that UCAB graduates in computer engineering, must achieve to their professional practice. Finally, with this first version of the skill units of computer engineer, we define general engineer skills for faculty careers in telecommunications, industrial and civil.

Keywords: Curricula Reform; Tuning Project; EEES; Computer Engineering.

I. INTRODUCCIÓN

Los acuerdos europeos, iniciados a partir de la declaración de Bologna el 19 de junio de 1999, lanzan la iniciativa del EEES (Espacio Europeo de Educación Superior). La intención es definir pregrados, maestrías y doctorados con múltiples puntos en común que permitan sistemas de titulación comparables, pensum de estudio más flexibles, mayor movilidad estudiantil, introducción de las tecnologías de la información en la enseñanza, etc [1]. En efecto, las exigencias de las empresas y de la sociedad del conocimiento exigen una reestructuración radical donde la enseñanza se centra en las capacidades críticas del estudiante que lo impulsen al autoaprendizaje. Entre las habilidades a desarrollar, debe

manejar la heterogeneidad en grupos de profesionales donde el trabajo en equipo y las capacidades de liderazgo son de suma importancia para lograr objetivos comunes.

Este movimiento de cambio en las universidades europeas se ha expandido a nivel mundial avalados con el proyecto Tuning [2], que se enfoca en Europa y se ha diversificado hacia América Latina. La idea inicial del proyecto es poner en concordancia las universidades europeas en términos de competencias, criterios de desempeño y resultados de aprendizaje. La intención no es imponer ningún modelo o estrategia sino compartir los puntos comunes y "sintonizar" las universidades participantes en la búsqueda de una universidad que se adapte a la sociedad donde ejercen sus

egresados. En un principio el proyecto fue puesto en marcha, justo después de la declaración de Bologna, por las Universidades de Deusto y de Groningen y ha continuado expandiéndose más allá de las fronteras europeas.

Para definir las competencias de sus egresados, algunas instituciones de educación superior se han concentrado en desarrollar sus capacidades de e-learning y educación a distancia, otras se han centrado en la adecuación con su entorno empresarial para responder a lo que requiere el país, aún otras en el desarrollo de las habilidades críticas necesarias para el manejo de la gran cantidad de información que deben afrontar los estudiantes. Efectivamente, todos estos aspectos son relevantes para la definición de un plan de estudios basado en competencias.

Justamente las competencias son un término clave para esta reforma curricular. La filosofía de transformación en las universidades gira alrededor de este concepto cuyo significado en español puede generar confusión pues es polisémico. Competencia se puede interpretar como una disputa entre distintas personas para cumplir objetivos, muy asociado a eventos deportivos. Por otro lado puede interpretarse como incumbencia o grado de compromiso de una persona con respecto a un asunto. Sin embargo el concepto más cercano a los pensum de estudio es el que se asocia con pericia o destreza ante un asunto o actividad. Para ciertos autores la interpretación correcta es una mezcla de estos diferentes significados ... en todo caso lo que resulta evidente es que el desarrollo de carreras por competencias intentan dar respuesta a las exigencias que plantea la nueva sociedad del conocimiento.

En un primer momento las competencias se dividen en genéricas o transversales y técnicas o específicas. Las transversales pueden ser definidas para todo egresado de una universidad o comunes a distintas carreras pertenecientes a la misma facultad. Las específicas son más técnicas, por ejemplo particulares de cada carrera, generalmente agrupada por dimensiones extraídas de un perfil del egresado. Más detalladamente, las competencias también pueden clasificarse según tres grandes bloques: disciplinares (saber), instrumentales (saber hacer) y actitudinales (ser). Los dos primeros tipos están íntimamente vinculados al conocimiento particular del área de conocimiento. En cambio las competencias actitudinales se relacionan más, en cierto modo, con propiedades personales que se espera desarrolle el estudiante.

Para facilitar la tarea de llevar una reforma por competencias, además del proyecto Tuning, las universidades españolas definieron documentos de procedimientos para cada carrera conocidos como los libros blancos auspiciado por la Agencia Nacional de Evaluación de la Calidad y Acreditación [3]. Así, junto con el proyecto Tuning, constituyen una valiosa herramienta para la transformación universitaria avalado por aquellas universidades que ya han comenzado y terminado el proceso de reforma. En este trabajo nos apoyamos en ambos documentos y varios artículos con resultados y técnicas que han utilizado diversas universidades en sus procesos de reforma curricular.

Por último están los trabajos de la ACM (Association for Computing Machinery) sobre perfiles curriculares específicos a la computación [4]. Se trata de una asociación norteamericana que agrupa profesionales, docentes y científicos cuyo objetivo es la difusión del conocimiento y patrocinio de eventos del área de computación y afines. Tiene diversas actividades y proyectos como el maratón de programación, las conferencias y revistas especializadas, la biblioteca digital y las recomendaciones de reforma curricular desde la década de los 60. Este último proyecto reúne a docentes y profesionales, especialistas, en diferentes subáreas de la computación, quienes ante los permanentes avances tecnológicos, mantienen actualizado una propuesta de currícula en computación en diferentes subcarreras. Aunque no nació con la misma motivación que la EEES tienen una estrecha vinculación con la acreditación y reforma curricular necesaria para fijar puntos comunes que permitan comparar las diferentes carreras afines a computación. En esta primera fase del trabajo no se consideran estas recomendaciones aunque son parte de trabajos futuros.

Así el objetivo de este trabajo, en su primera fase, fue definir las competencias para el ingeniero en informática a partir de las competencias genéricas definidas para todo egresado de la Universidad Católica Andrés Bello. Aplicando un análisis factorial sobre ciertas variables medidas mediante una encuesta, se ajustó el perfil del egresado en Ingeniería Informática. Finalmente, junto con las demás ingenierías de la facultad, coordinar la definición de las competencias del ingeniero de nuestra universidad.

II. TRABAJOS RELACIONADOS

Como ya se mencionó, el proyecto *Tuning* constituye nuestro punto de partida en la definición de las competencias genéricas del ingeniero en informática egresado de la Universidad Católica Andrés Bello. Las directivas partieron desde el equipo rectoral y fueron descendiendo hasta llegar a las instancias de especialistas en las diversas áreas para la puesta en marcha de la reforma curricular.

En algunas universidades se ha modificado su estructura interna para acoplarse a los cambios como la Universidad de Extremadura que creó un Vicerectorado de Docencia e Integración Académica que evolucionó al Vicerectorado de Planificación Académica [1].

La universidad de Oviedo [5] realizó un trabajo de reforma por competencias utilizando el libro blanco al cual agregó varias modificaciones (llamado por ellos el libro azul). Sin embargo el trabajo, específicamente para Ingeniería en Informática, fue por consenso entre un grupo de profesores (trece en total) mediante criterios cualitativo sin valerse de ningún instrumento de consulta que incluyese la población universitaria.

La facultad de informática de la Universidad Politécnica de Cataluña definió sus criterios [6] apoyándose en la metodología *tuning* [2] para Europa y usando los cinco perfiles definidos por ACM [4]:

- Computer Engineering
- Computer Science

- Information Systems
- Software Engineering
- Information and Communication Technology

En el artículo participaron cinco profesores de los departamentos de Arquitectura del Computador, Lenguajes y Sistemas Informáticos y Estadísticas e Investigación de Operaciones en aras de buscar generalidad en la definición de sus competencias. La aplicación de instrumentos de consulta está entre sus trabajos futuros.

También un grupo de investigadores chilenos, de la Universidad de Talca, proponen una metodología cuyo punto de partida es asociar a cada competencia un módulo o una asignatura, asociación que puede ser modificada por restricciones de tiempo y recursos a través de un proceso de refinamientos sucesivos. Aún no han publicado resultados sobre la aplicación de esta metodología [7].

Otras universidades han contratado equipo de asesores y formado docentes para coordinar el trabajo de transformación para toda la universidad con un enfoque *top-down*. Es decir definir las competencias del egresado de dicha universidad (competencias transversales) para hacerlo específico por cada facultad, escuela, coordinación y finalmente fijar las competencias por asignatura (competencias específicas). Este enfoque es el utilizado en la Universidad Católica Andrés Bello y será mostrado en este artículo, específicamente para la Escuela de Ingeniería Informática.

III. COMPETENCIAS GENERALES DE LA UNIVERSIDAD

Desde el punto de vista más general las competencias deberían ser una concatenación de saberes, no sólo pragmáticos y orientados a la producción, sino directivas que articulan una concepción del ser, del saber, del saber hacer y del saber convivir. Es por ello que la Universidad Católica Andrés Bello, gracias a un grupo de docentes formados en Deusto y asesores expertos, formuló las competencias que debe tener el egresado de la universidad. Estas fueron avaladas por las instancias y consejos que incluyen autoridades de todas las escuelas de pregrado y postgrado. Las competencias propuestas, articuladas bajo las directivas generales son:

A. *Aprender a Aprender con Calidad*

Utiliza estrategias de forma autónoma para incorporar e incrementar conocimientos, habilidades y destrezas en el contexto de los avances científicos y culturales requeridos para un ejercicio profesional globalmente competitivo.

Unidades de competencia:

- Abstrae, analiza y sintetiza información
- Aplica los conocimientos en la práctica
- Identifica, plantea y resuelve problemas
- Incorpora conocimientos y se actualiza permanentemente
- Se comunica eficazmente de forma oral y escrita
- Demuestra conocimiento sobre su área de estudio y profesión

- Realiza investigaciones
- Trabaja con altos estándares de calidad
- Busca y procesa información de diversas fuentes
- Trabaja en forma autónoma

B. *Aprender a Convivir y a Servir*

Reconoce, aprecia y cultiva de manera reflexiva, ética, responsable y comprometida, su relación con otras personas y con el medio ambiente físico y sociocultural, local y global, para contribuir al bienestar colectivo.

Unidades de competencia:

- Se involucra con su medio socio-cultural
- Valora y respeta la diversidad y multiculturalidad
- Participa y se involucra en actividades comunitarias y ciudadanas
- Participa activamente en la preservación del medio ambiente
- Reflexiona y cuestiona su propia actuación
- Actúa según valores éticos compartidos
- Se solidariza con el otro

C. *Aprender a Trabajar con el Otro*

Interactúa con otros en situaciones diversas y complejas para alcanzar objetivos comunes, en un entorno donde el equilibrio de los roles: colaborador o líder y la fluidez comunicativa procuran resultados beneficiosos para todos.

Unidades de competencia:

- Participa y trabaja en equipo
- Motiva y conduce a otros hacia metas comunes
- Toma decisiones efectivas para resolver problemas
- Actúa eficazmente en nuevas situaciones
- Colabora, interactúa, se comunica con el otro
- Organiza y planifica el tiempo
- Formula y gestiona proyectos
- Plantea y resuelve problemas creativamente

D. *Aprender a Interactuar en el Contexto Global*

Actúa y se integra a los escenarios globales mediante el dominio de otros idiomas y de las tecnologías de la información y comunicación, esenciales para su interacción en el escenario global.

Unidades de Competencia:

- Se desempeña eficazmente en contextos internacionales.
- Maneja adecuadamente las tecnologías de información y comunicación
- Se comunica con fluidez en un segundo idioma

IV. COMPETENCIAS DE INGENIERÍA INFORMÁTICA

Una vez definidas las competencias generales del egresado de la UCAB, se inició un proceso de reforma curricular por

competencias en todas las escuelas de pregrado. Para ese momento, en el año 2008, la escuela de Ingeniería Informática, se encontraba en pleno cambio de pensum ajustado a los avances académicos, algunas necesidades de los empleadores y ciertas consideraciones de la población estudiantil. Este proceso se realiza periódicamente para garantizar la actualización de los contenidos, con base a pensum de estudios de carreras afines tanto nacionales como internacionales. Así se puede mantener al egresado en la vanguardia tecnológica y también responder a ciertas necesidades de los empleadores.

En ese momento, se decidió finalizar con ese cambio de pensum y evaluar otras ideas como pasar a períodos trimestrales en cuatro años. Para organizar mejor las actividades se decidió concluir esa revisión sin contemplar la reforma curricular por competencias, ni el cambio a períodos trimestrales en cuatro años.

Una vez actualizado los planes de estudio, correspondía adaptar los programas ya revisados bajo los lineamientos de la reforma por competencias. Para ello se comienza un plan para la definición de las competencias específicas del ingeniero en informática. Se revisan los pensum de estudios orientados por competencias de otras Universidades de habla hispana, particularmente el de la Universidad de Deusto, la cual, como la UCAB, forma parte de las universidades adscritas a la compañía de Jesús.

El proceso realizado para lograr estas competencias específicas fue:

1. Se reescribió el perfil del ingeniero en informática incorporando dimensiones que se creía deberían abarcar las competencias profesionales, el perfil quedó definido así:

El Ingeniero en Informática de la Universidad Católica Andrés Bello es un profesional emprendedor con una formación integral que brinda soluciones efectivas a problemas relacionados con el diseño, producción e implantación de software de aplicación y sistemas telemáticos; administra y mantiene sistemas operativos y equipos de computación; planifica, dirige y controla proyectos informáticos, apoyándose en herramientas de toma de decisiones.

2. Cada coordinación académica de la Escuela de Ingeniería Informática propuso una lista de competencias con base a los objetivos de las materias correspondientes a esta coordinación. Esta lista fue trabajada en conjunto con los profesores de dichas materias.
3. Las listas fueron revisadas y depuradas entre todos los profesores tiempo completo (ocho docentes donde además la mayoría son coordinadores de área académica). Esta depuración consistió en eliminar competencias repetidas, refinar redacción y llegar a un consenso general. Se obtuvo una lista preliminar de 24 competencias profesionales que se listan a continuación

- 3.1. Aplica diferentes paradigmas de análisis, diseño y programación para implementar software de aplicación.
- 3.2. Elabora toda la documentación técnica y funcional de un sistema informático.
- 3.3. Diseña y ejecuta planes de pruebas de los sistemas informáticos.
- 3.4. Utiliza la abstracción para diseñar e implementar en forma modular y sistemática una solución algorítmica eficiente.
- 3.5. Resuelve un problema en forma lógica, analítica y estructurada argumentando correctamente cualquier solución que se plantee en el área de diseño de software.
- 3.6. Diseña e implementa software de aplicación sobre arquitecturas centralizadas y distribuidas
- 3.7. Planifica y controla la ejecución de un proyecto de Tecnología de Información
- 3.8. Identifica, selecciona y aplica la metodología apropiada para el desarrollo de software de aplicación
- 3.9. Diseña, implementa y administra un sistema de Bases de Datos
- 3.10. Escribe programas eficientes conociendo el funcionamiento y organización del computador.
- 3.11. Representa programas a nivel máquina.
- 3.12. Administra y mantiene sistemas operativos basado en los conceptos y funcionamiento.
- 3.13. Utiliza la abstracción para identificar el modelo matemático que se adapte a una situación real.
- 3.14. Formula, resuelve y analiza modelos matemáticos, que representen una situación real, para apoyar la toma de decisiones.
- 3.15. Representa y simplifica un problema mediante la construcción de un modelo matemático, esto permite analizar con claridad un objeto de estudio eliminando lo superfluo.
- 3.16. Diseña y construye sistemas electrónicos digitales.
- 3.17. Programa microcontroladores.
- 3.18. Diseña y construye dispositivos digitales que pueden interactuar con un computador personal.
- 3.19. Diseña, instala y administra redes por medios de transmisión física e inalámbrica
- 3.20. Concibe e implanta mecanismos de seguridad en sistemas informáticos.
- 3.21. Planifica y desarrolla un proceso de evaluación integral de un sistema de informático.
- 3.22. Evalúa la factibilidad técnica y económica de un proyecto informático para obtener la mejor rentabilidad.
- 3.23. Vincula la ética con el ejercicio profesional
- 3.24. Se comunica con eficacia y propiedad de forma oral y escrita

4. Se consultó a una población asociada a la Escuela de Ingeniería Informática, mediante una encuesta, sobre la importancia de cada competencia para el perfil del egresado y el grado de desarrollo que logran los estudiantes al final de su escolaridad. El instrumento se aplicó a 25 profesores, 30 empleadores, 57 egresados y 51 estudiantes de los últimos semestres.

Tabla I: Resultados de Importancia y Desarrollo de la Consulta a Profesores, Empleadores, Egresados y Estudiantes de Último Año para cada Competencia (Ordenado por Orden Descendente de Importancia)

Competencia	Importancia	Desarrollo
5. Resuelve un problema en forma lógica, analítica y estructurada argumentando correctamente cualquier solución que se plantee en el área de diseño de software.	3,74	3,17
24. Se comunica con eficacia y propiedad de forma oral y escrita	3,73	2,80
9. Diseña, implementa y administra un sistema de Bases de Datos	3,69	3,51
23. Vincula la ética con el ejercicio profesional	3,57	3,13
4. Utiliza la abstracción para diseñar e implementar en forma modular y sistemática una solución algorítmica eficiente.	3,55	3,09
7. Planifica y controla la ejecución de un proyecto de Tecnología de Información	3,53	2,61
8. Identifica, selecciona y aplica la metodología apropiada para el desarrollo de software de aplicación	3,51	3,02
1. Aplica diferentes paradigmas de análisis, diseño y programación para implementar software de aplicación	3,49	3,05
6. Diseña e implementa software de aplicación sobre arquitecturas centralizadas y distribuidas	3,43	2,80
22. Evalúa la factibilidad técnica y económica de un proyecto informático para obtener la mejor rentabilidad.	3,42	2,48
20. Concibe e implanta mecanismos de seguridad en sistemas informáticos.	3,40	2,47
21. Planifica y desarrolla un proceso de evaluación integral de un sistema de	3,40	2,47

informático.		
3. Diseña y ejecuta planes de pruebas de los sistemas informáticos.	3,38	2,43
2. Elabora toda la documentación técnica y funcional de un sistema informático	3,29	2,83
10. Escribe programas eficientes conociendo el funcionamiento y organización del computador.	3,13	2,66
19. Diseña, instala y administra redes por medios de transmisión física e inalámbrica	3,02	2,40
14. Formula, resuelve y analiza modelos matemáticos, que representen una situación real, para apoyar la toma de decisiones.	2,86	2,42
12. Administra y mantiene sistemas operativos basado en los conceptos y funcionamiento.	2,76	2,20
15. Representa y simplifica un problema mediante la construcción de un modelo matemático, esto permite analizar con claridad un objeto de estudio eliminando lo superfluo.	2,67	2,30
13. Utiliza la abstracción para identificar el modelo matemático que se adapte a una situación real.	2,64	2,41
18. Diseña y construye dispositivos digitales que pueden interactuar con un computador personal.	2,46	2,11
16. Diseña y construye sistemas electrónicos digitales.	2,27	2,22
17. Programa microcontroladores.	2,15	2,11
11. Representa programas a nivel máquina.	2,09	2,11

5. Con la tabla anterior se realizó un análisis factorial para detectar agrupamiento de las competencias por características no indicadas explícitamente. Esas características se transformaron en las dimensiones del egresado que corroboraron la primera aproximación del perfil del ingeniero informático de la UCAB.
6. Para cada una de las cuatro dimensiones obtenidas a partir del análisis factorial, se definieron las respectivas unidades de competencia, indicando para cada una de ellas, los criterios de desempeño. Esta definición se realizó a través

de reuniones con los profesores a tiempo completo de la escuela y en algunos casos, se invitó a profesores convencionales (contratados por horas) de las áreas profesionales de la carrera. Las dimensiones unidades de competencia se estructuran así:

- 6.1. Gestionar Proyectos Informáticos: planifica, dirige y evalúa proyectos informáticos de forma efectiva en distintos ambientes organizacionales, participando activamente en equipos de trabajo interdisciplinarios, orientados a la búsqueda de soluciones con sentido ético.

Unidades de competencia:

- Planifica y desarrolla un proceso de evaluación integral de un sistema informático.
- Planifica y controla el desarrollo de un proyecto de tecnología de Información
- Documenta técnica y funcionalmente un sistema informático

- 6.2. Emprender Negocios de Ingeniería: identifica la oportunidad de negocio, crea la empresa y desarrolla el negocio.

Unidades de competencia:

- Identifica oportunidades de negocio
- Ejecuta el plan de negocio

- 6.3. Desarrollar Sistemas Telemáticos: diseña, implementa, construye circuitos electrónicos y sistemas de comunicación seguros y confiables en distintos ambientes operativos.

Unidades de competencia:

- Diseña y construye dispositivos electrónicos
- Diseña, instala y administra redes de voz y datos
- Diseña e implementa sistemas de seguridad en redes
- Administra y mantiene sistemas operativos

- 6.4. Desarrollar Software de Aplicación: identifica, analiza, diseña, implementa, evalúa y mantiene software de aplicación de manera efectiva.

Unidades de competencia:

- Diseña e implementa algoritmos robustos y eficientes de forma modular
- Analiza las necesidades de los usuarios, diseña e implementa el software de aplicación sobre arquitecturas centralizadas o distribuidas
- Diseña y ejecuta planes de pruebas de los sistemas informáticos
- Identifica, selecciona y aplica la metodología apropiada para el desarrollo de software de aplicación
- Diseña, implementa y administra un Sistema de Base de Datos.

Desde la facultad de ingeniería, se realizó el trabajo de definición de las competencias generales del ingeniero de la UCAB, con las Escuelas de Ingeniería: Industrial, Civil y Telecomunicaciones, inspirados en el trabajo realizado por la Escuela de Ingeniería Informática. Las actividades fueron:

- Se reunieron las comisiones de currículo de todas las escuelas de la Facultad de Ingeniería tanto de su sede en Caracas como las de Guayana.
- Cada escuela llevó una propuesta sobre estas competencias básicas, y por consenso se determinaron cuales serían las dimensiones generales con sus respectivos criterios de evaluación.
- Bajo la coordinación de la Escuela de Ingeniería Informática, se definieron las competencias generales del ingeniero. Al finalizar este proceso contaremos con la malla curricular basada en competencias que debe ser validada, como se hizo en la Escuela de Ingeniería Informática, por los profesores (ciclo profesional y básico), empleadores, egresados y estudiantes del último año.

A continuación se indican las competencias básicas del Ingeniero de la UCAB (Informática, Civil, Industrial y Telecomunicaciones) avaladas por los directores de escuela y asesores del proyecto de reforma curricular por competencia.

V. COMPETENCIAS BÁSICAS DEL INGENIERO DE LA UCAB

A. Formular Proyectos de Ingeniería

Formula, planifica y evalúa proyectos de ingeniería de forma efectiva en distintos ambientes organizacionales, participando activamente en equipos de trabajo interdisciplinarios, orientados a la búsqueda de soluciones con sentido ético.

Unidades de competencia:

- Evalúa la factibilidad técnica y económica de un proyecto de ingeniería
- Cumple con el código de ética profesional y el marco legal vigente

B. Modelar para la Toma de Decisiones

Representa a través de modelos matemáticos situaciones de la vida real, para apoyar de forma efectiva la toma de decisiones en base a un objetivo planteado.

Unidades de competencia:

- Modela matemáticamente situaciones reales para apoyar la toma de decisiones
- Simula computacionalmente situaciones de la vida real

VI. CONCLUSIONES Y TRABAJOS FUTUROS

De los resultados de la encuesta sobre competencias generales de los ingenieros en informática se aprecia que competencias transversales como ética y expresarse correctamente en forma oral y escrita quedaron con un alto nivel de importancia para los encuestados. A nuestro juicio estas competencias se acercan a desarrollar habilidades básicas, que los encuestados consideramos primordiales para la actual sociedad

venezolana. Por otro lado competencias más técnicas como representar problemas con modelos matemáticos que parecen fundamentales para el desarrollo de soluciones informáticas, tiene para los encuestados un bajo nivel de importancia. Seguramente las consideraciones prácticas son más relevantes para la población universitaria y empleadores, aún en detrimento de aspectos teóricos. Por otro lado el nivel de importancia y desarrollo de áreas como bases de datos y metodologías para el desarrollo de software están entre los más altos. Estos resultados los avala el estudio de egresados realizado por la Escuela de Ingeniería Informática [8] y que se considera entre los puntos fuertes de nuestros egresados.

Actualmente todos los docentes de la escuela de ingeniería informática, están validando cada una de las dimensiones, junto con las unidades de competencia, mediante una encuesta por materias existentes en el pensum actual. En esta fase se espera precisar las competencias según la opinión de la población encuestada y entonar nuestras materias. Además los profesores a tiempo completo se están formando en el Diplomado de Docencia Universitaria Orientada al Desarrollo de Competencias compuesto de tres módulos: Identidad Ucabista, Competencias Docentes de Profesor Universitario y Estrategias Didácticas y Evaluación de Competencias. Este módulo de formación es primordial para la reforma curricular por competencia y no es el único necesario. Dentro del plan de formación integral de la UCAB están previstos otros módulos como utilización de las TIC en la enseñanza que nos permitirán consolidar esta reforma.

Como trabajo a futuro, vamos a comparar nuestro pensum de estudio con los perfiles de carreras afines a informática definido por ACM para constatar a cual pertenecemos. También validar este trabajo contra las competencias específicas del libro blanco para ingeniería informática.

REFERENCIAS

- [1] M. Montero, *El Proceso de Bolonia y las Nuevas Competencias*, Revista Tejuelo, no. 9, pp. 19-37, Extremadura, España, 2010.
- [2] J. Gonzalez y R. Wagenaar, *Tuning Educational Structure in Europe*, Informe Final, Fase Uno, Reporte interno de la Universidad de Deusto, <http://www.unideusto.org/tuning>, 2003.
- [3] J. Casanovas, J. Colom, I. Morlán, A. Pont y M. Snacho, *El libro Blanco de la Ingeniería en Informática: el Proyecto EICE*, http://www.aneca.es/var/media/150388/libroblanco_jun05_informatica.pdf.
- [4] IEEE/ACM, *Curricula Recommendations*, <http://www.acm.org/education/curricula.html>.
- [5] A. Aquilino, J. de Andrés, C. Nieto, et al, *Definición y Competencias Específicas y Genéricas del Ingeniero en Informática*, Docencia Universitaria y Proyectos de Innovación Docente, 2006.
- [6] F. Sánchez, M-R. Sancho, P. Botella, J. García, T. Aluja, J. Navarro y J. Balcazar, *Competencias Profesionales del Grado en Ingeniería Informática*, Jornadas de Enseñanza Universitaria de la Informática (JENUI 2008), Granada, España.
- [7] R. Schmal, A. Ruiz-Tagle, *Una Metodología para el Diseño de un Currículo Orientado a las Competencias*, *Ingeniare, Revista Chilena de Ingeniería*, vol. 16, no. 2, pp. 147-158, 2008.
- [8] R. Paladino, *Estudio y Análisis de Encuestas sobre Egresados de la UCAB*, Reporte interno de Universidad Católica Andrés Bello, Enero 2013.

Un Framework para la Rehabilitación Física en Miembros Superiores con Realidad Virtual

Esmitt Ramírez¹, Francisco Moreno¹, Jordan Ojeda¹, Christiam Mena¹,
Omaira Rodríguez^{1,2}, Jeanlight Rangel³, Silvio Álvarez³

esmitt.ramirez@ciens.ucv.ve, morenof90@gmail.com, jrbworr@gmail.com, christiam.mena@gmail.com,
omaira.rodriguez@gmail.com, jeanlight.rangel2@gmail.com, slab_ve@hotmail.com

¹ Centro de Computación Gráfica, Facultad de Ciencias, Universidad Central de Venezuela. Caracas, Venezuela

² Instituto Nacional de Bioingeniería, Universidad Central de Venezuela. Caracas, Venezuela

³ Servicio de Medicina Física y Rehabilitación, Hospital de Niños J.M. de Los Ríos. Caracas, Venezuela

Resumen: La rehabilitación física es una herramienta médica para el re-entrenamiento de personas afectadas por lesiones adquiridas o enfermedades que pueden afectar al sistema nervioso dando como consecuencia un mal funcionamiento del control motor voluntario. Recientemente, han surgido vertientes de gran impacto en términos de realidad virtual asociadas a la rehabilitación física de pacientes demostrando ser herramientas efectivas de ayuda terapéutica y diagnóstica. En muchas ocasiones, los resultados obtenidos con la realidad virtual superan a los obtenidos con metodologías tradicionales. En este trabajo se presenta un framework para la rehabilitación física para miembros superiores en pacientes entre 6 y 12 años de edad basado en realidad virtual. Nuestra propuesta se fundamenta en la creación de una arquitectura de hardware y software, donde se ejecutan un conjunto de videojuegos que inducen actividades físicas particulares, y a la provisión de cambios neuronales mediante el empleo de la acción/observación. Además, se procura la captura de la atención del paciente empleando contenido dinámico y entretenido de forma interactiva. El framework fue diseñado en conjunto con personal especializado que coordina la terapia de rehabilitación del Hospital de Niños J.M. de Los Ríos ubicado en Caracas, Venezuela.

Palabras Clave: Realidad Virtual; Terapia de Rehabilitación; Videojuegos; Activación Cerebral; Kinect.

Abstract: Physical rehabilitation is a medical tool for re-training of affected people caused by injuries or diseases that can affect the nervous system giving as result the malfunction of the voluntary motor control. Nowadays, there have been aspects of great impact in terms of virtual reality associated with physical rehabilitation of patients proving to be effective tools for diagnostic and therapeutic aid. In many cases, results obtained using virtual reality are better than those obtained with traditional methodologies. This paper presents a computational framework for physical rehabilitation for upper limbs in patients from 6 to 12 years old based on virtual reality. Our proposal is based on the creation of a hardware and software architecture, where a set of games that induce particular physical activities, and the provision of neuronal changes by using the scheme of action/observation. Also, we seek the complete attention of patients with the use of dynamic and entertaining content interactively. The framework was designed together with a specialized staff to coordinate the rehabilitation therapy at Hospital de Niños J.M. de Los Ríos (a children's hospital) located in Caracas, Venezuela.

Keywords: Virtual Reality; Rehabilitation Therapy; Video Games; Cerebral Activation; Kinect.

I. INTRODUCCIÓN

En el año 2011 la Organización Mundial de la Salud (OMS) estimó que más de 1.000 millones de personas en el mundo experimentan algún tipo de discapacidad [1]. En Latinoamérica más del 25% de la población con discapacidades son niños y adolescentes, y solo del 2% a 3% de esta población tiene acceso a los programas y servicios de rehabilitación [2].

Son muchas las formas de expresión, alcances y consecuencias

que acarrear las discapacidades. Así, una parte de la población afectada evidencia distintos tipos de déficit sensitivos, cognitivos y motores. Particularmente, relacionado con los déficit motores existen diversos tipos de discapacidad que afectan distintas partes del cuerpo con distintos grados de severidad.

La rehabilitación física es el conjunto de medidas sociales, educativas y profesionales que concierne a la evaluación, diagnóstico, prevención y tratamiento de la incapacidad en-



Figura 1: Fotografías del Ambiente Virtual Creado para la Rehabilitación Física en Miembros Superiores

caminados a facilitar, mantener o devolver el mayor grado de capacidad funcional e independencia posible.

El avance de las tecnologías ha permitido el desarrollo de procedimientos que permiten explorar la actividad cerebral durante las terapias de rehabilitación. Donde la actividad cerebral constituye un factor de vital importancia en la rehabilitación motora. Recientemente, con el fin de determinar las técnicas más efectivas y eficientes, una vertiente de estudios se ha orientado a tratar distintas metodologías de re-entrenamiento tomando en cuenta este aspecto.

Actualmente la realidad virtual es altamente empleada en rehabilitación, generando investigaciones y aplicaciones que contribuyen en esta rama de la medicina. La terapia basada en realidad virtual cuenta con un contexto funcional, concreto y estimulante para los pacientes, trayendo un beneficio directo tanto al paciente como al terapeuta por la adaptabilidad que tienen estos sistemas. Las tecnologías de punta son empleadas para la producción de ambientes simulados interactivos y multidimensionales. Los dispositivos visuales como monitores y lentes, dispositivos hápticos, hardware para el seguimiento corporal, etc. son introducidos para sumergir al paciente en un entorno virtual y dotarlo de la capacidad de modificarlo en función de metas a cumplir.

En la búsqueda de alternativas para la aplicación de realidad virtual en procura de mayores beneficios para pacientes tratados con estas tecnologías, se incorporan dispositivos que inicialmente han sido diseñados para otros fines pero que han mostrado ser funcionales en esta área, generando su adopción en estas terapias como el Xbox 360[®] Kinect[®], Wii[®], etc. Adicionalmente, su bajo costo en comparación con hardware especializado lo hace aún más atractivo.

En este trabajo se presenta un *framework* para el desarrollo de un sistema de realidad virtual no inmersiva con el fin de construir soluciones en el área de rehabilitación física empleando el hardware del Kinect[®]. Nuestra propuesta plantea el uso de una arquitectura para un sistema de bajo costo basado en videojuegos para la rehabilitación motora, enfocado en la activación y re-entrenamiento de miembros superiores, dirigido a niños entre 6 y 12 años de edad, ver Figura 1. En este trabajo, nos referiremos al término *pacientes* para indicar este rango de edad. Empleando el *framework* es posible aplicar, extender y ejecutar diversos videojuegos sobre un

sistema controlado por un especialista o terapeuta.

Adicionalmente, es importante destacar que esta investigación está enmarcada en un proyecto interdisciplinario en conjunto con el Hospital de Niños J.M. de Los Ríos de Caracas cuyos expertos en el área de Rehabilitación y Medicina Física ajustan los requerimientos del sistema para su uso clínico en los pacientes

En la Sección II se muestra una pequeña introducción a los conceptos asociados al uso de la realidad virtual en la rehabilitación. La Sección III presenta una revisión de los trabajos previos aplicados al tema de estudio. Luego, la Sección IV explica detalladamente la propuesta de este trabajo. En la Sección V, se muestra los aspectos del desarrollo de la plataforma de rehabilitación así como el empleo del *framework*. La Sección VI presenta una evaluación del sistema a nivel de software. Finalmente, en la Sección VII se muestra una discusión y conclusiones de los aspectos de nuestro trabajo.

II. LA REALIDAD VIRTUAL EN LA REHABILITACIÓN

El término realidad virtual (RV), fue popularizado en el año 1986 por Jaron Lanier, desde entonces, ha sufrido numerosas transformaciones, debido al constante avance de la tecnología. Actualmente se puede definir como una simulación del mundo real generada por un software y experimentada mediante una interfaz humano-computador.

La realidad virtual se clasifica de acuerdo al grado de inmersión que genera en el usuario. Según Burdea y Coiffet [3], el término inmersión es el acto voluntario de obviar los distintos estímulos que hace percibir la experiencia presentada como no real, permitiendo captar toda la atención y concentración del usuario involucrado. Así, resulta posible clasificar la RV en dos vertientes, inmersiva y no inmersiva. La primera envuelve la mayor cantidad de sentidos posibles del usuario (generalmente empleando hardware especializado). La segunda, no inmersiva, es más común (no requiere de hardware especializado) y se encuentra a menudo en los videojuegos.

En el área de la medicina y psicología, la RV ha sido ampliamente utilizada, dado que puede presentar situaciones desafiantes dentro de un ambiente seguro, manteniendo un control experimental sobre la medición y presentación de los estímulos [4]. Las primeras aplicaciones de este tipo dieron

lugar a una serie de intervenciones psicosociales para el tratamiento de fobias. Luego, se expandió a otras áreas como entrenamiento quirúrgico y rehabilitación motora, entre otras.

En la rehabilitación motora, la RV se emplea como una herramienta de rehabilitación y evaluación en programas de entrenamiento del equilibrio, la postura y la marcha, activación de las funciones en los miembros superiores, terapias de tolerancia al ejercicio y al dolor, evaluación de actividades de la vida diaria y evaluación de negligencias visuales posteriores a un accidente cerebro vascular (ACV) [5].

De acuerdo a Holden [6], en un proceso de rehabilitación motora exitoso se destacan tres aspectos fundamentales: repetición, motivación y retroalimentación. Para aprender a realizar un movimiento, este debe ser ejecutado repetidamente para la memorización y el aprendizaje motor, ya que estimula la reorganización neuronal. La motivación es un aspecto donde la terapia a través de RV suele destacarse, debido a la forma como se puede presentar un ejercicio, resultando agradable para el usuario. En un videojuego se plantean distintas metas a cumplir, las cuales dan una retroalimentación al usuario de estar realizando un movimiento de forma correcta o no. Si tales objetivos son planteados correctamente entonces representan un reto al usuario, generando una motivación para lograr el objetivo. Esto permite realizar las repeticiones necesarias para estimular la reorganización cerebral y así, memorizar el movimiento.

Los ejercicios terapéuticos propuestos usando RV suelen ser totalmente configurables y, en los ambientes virtuales (AV) generados se da especial atención a los puntos clave del ejercicio, eliminando los posibles agentes distractores que puedan degradar la calidad de la terapia. A continuación, se estudiarán algunos de los estudios realizados en esta dirección.

III. TRABAJOS PREVIOS

Los primeros trabajos de RV para rehabilitación terapéutica fueron bajo estudios con pacientes sanos, y posteriormente aplicados a pacientes con alguna discapacidad. Webster et al. [7] realizaron un experimento con pacientes afectados por ACV en sillas de ruedas y con el síndrome de negligencia. En su experimento, se llevaron a cabo pruebas empleando terapia tradicional y terapia de RV, observando menos errores en el recorrido de una trayectoria con obstáculos al usar RV.

En años recientes los AV han sido utilizados como estudios experimentales aportando buenos resultados. Merians et al. [8] plantean la posibilidad que los efectos de la plasticidad estimulados con terapias basadas en RV conducen a resultados positivos. Al mismo tiempo, sugieren que la rehabilitación virtual (término acuñado a este proceso) debe basarse en la comprensión de la potencialidad del sistema nervioso para generar una modificación en la disposición neuronal. En el 2005, You et al. [9] demuestran que la terapia con RV produce cambios en la corteza sensorial-motora del cerebro, las cuales se relacionan con la mejora de las habilidades motrices de las extremidades afectadas por enfermedades causantes de la discapacidad.

BioTrak[®] desarrollado por Bienetec [10], es una herramienta basada en tecnología RV que integra en una misma plataforma ejercicios para el entrenamiento y rehabilitación de determinadas funciones que se han visto reducidas o perdidas a causa de diversas patologías. Igualmente, el sistema RGS [11] consiste en una serie de soluciones para rehabilitación de miembros superiores que contiene ambientes diseñados a la medida para cada paciente de forma configurable.

En los últimos años, las investigaciones se enfocan en utilizar hardware existente no especializado de bajo costo con el fin de masificar su utilización y ser más accesible. Un ejemplo de lo constituye el uso de hardware diseñado para el entretenimiento, dentro de la rehabilitación (e.g. Kinect[®] y Wii[®]).

Un ejemplo enfocado en la rehabilitación motora empleando Kinect[®] se presenta en el proyecto Kinesthesia [12], en esta aplicación el Kinect[®] captura la posición del cuerpo basado en puntos claves o uniones (joints). En 2012, LaBelle [13] menciona que el uso del Kinect[®] en la rehabilitación física es un tema abierto a la investigación dentro del área de la rehabilitación con mucho potencial en las terapias.

Resulta entonces factible estudiar y desarrollar soluciones haciendo uso de la tecnología Kinect[®] en términos de rehabilitación motora para constatar su inclusión en terapias que permitan obtener un proceso de recuperación exitoso. En la siguiente sección, mostramos nuestra solución basada en una arquitectura para rehabilitación física de miembros superiores empleando realidad virtual no inmersiva.

IV. SOLUCIÓN PROPUESTA

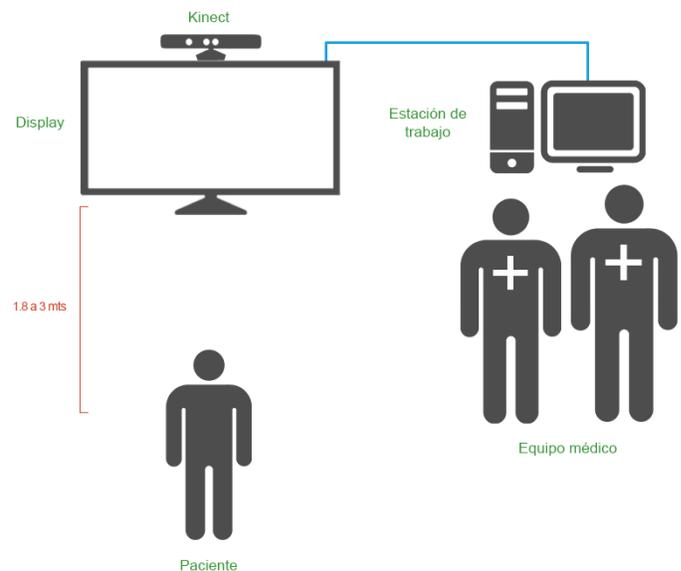


Figura 2: Arquitectura de Hardware Propuesta para nuestra Solución Terapéutica

Nuestra propuesta se basa en una arquitectura de hardware y software para la rehabilitación física de miembros superiores empleando realidad virtual que satisfaga un conjunto de requerimientos tales como, fácil de usar por el personal

médico, intuitivo y natural para el paciente, factible y seguro para el paciente, además de bajo costo y fácil de reproducir en instalaciones hospitalarias o el hogar. Nuestra propuesta ataca las actividades terapéuticas enfocadas en los miembros superiores.

La arquitectura de hardware propuesta consiste en una estación de trabajo con una PC convencional, un Kinect®, y un monitor LCD (display) tal como se muestra en la Figura 2.

El dispositivo Kinect®, detecta la posición del paciente, dotando al sistema de la capacidad de reconocimiento y reacción en función a la movilidad y gestos ejecutados por parte del mismo. El paciente se debe ubicar a una distancia de 1.8 m a 3 m del receptor Kinect®. Inicialmente el sistema realiza una calibración para determinar la posición y alcance de movimiento con los miembros superiores del paciente.

La arquitectura de software propuesta está conformada por dos componentes principales: uno dirigido al personal médico y otro a los pacientes. El primero de estos se encarga de la gestión administrativa del sistema, incluyendo el control de pacientes, sesiones y parámetros de los videojuegos. El segundo, está dedicado al despliegue del contenido gráfico con el que interactúan los pacientes, mostrando la ejecución de cada uno de los videojuegos.

A. Componentes de la Arquitectura

El componente desarrollado para el personal médico se muestra como una interfaz simple, intuitiva y fácil de usar. Ésta permite navegar e iniciar cada una de las opciones que el sistema provee. En la Figura 3 se muestran la vista al usuario de los cuatro módulos que forman este componente.



Figura 3: Interfaz del Componente Desarrollado para el Personal Médico

El primero de ellos consiste en la **Gestión-Paciente**, donde se muestra las opciones asociadas a la gestión de pacientes en la base de datos del sistema, que consiste en añadir a un nuevo paciente para realizar una terapia con el sistema, eliminar e incluso actualizar los datos de un registro. El módulo de **Calibración** incluye las opciones relativas a la activación de los métodos de calibración para el Kinect® que consiste en determinar los rangos de alcance motor de un paciente.

Luego, el módulo de **Mi Perfil** permite dar inicio a la sesión de un paciente y visualizar las estadísticas asociadas a su

historial de rehabilitación durante un tiempo determinado. Por último, el módulo **Juegos** maneja la selección de videojuegos y parámetros como el tiempo y la dificultad, antes del inicio de cada uno.

Es importante destacar que el personal médico podrá controlar todos los aspectos de la aplicación desde la interfaz que el sistema expone. En tal sentido, el paciente siempre jugará bajo las especificaciones que el terapeuta plantee.

Por otro lado, el componente desarrollado para los pacientes está basado en una interfaz NUI (Natural User Interface) presentada para responder a los eventos que se les presente en las actividades. Empleando el Kinect®, es posible identificar la ubicación del paciente así como los miembros superiores e identificarlos en la pantalla de juego. En cada juego, el paciente podrá observar su silueta capturada por el Kinect® y procesada por el sistema, de forma tal de asociar sus movimientos con los movimientos realizados en el juego (ver ejemplo en la Figura 1).

B. Estructura del Framework

La estructura del *framework* desarrollado se representa en módulos funcionales los cuales se pueden observar en la Figura 4. A continuación se detalla cada uno de estos.

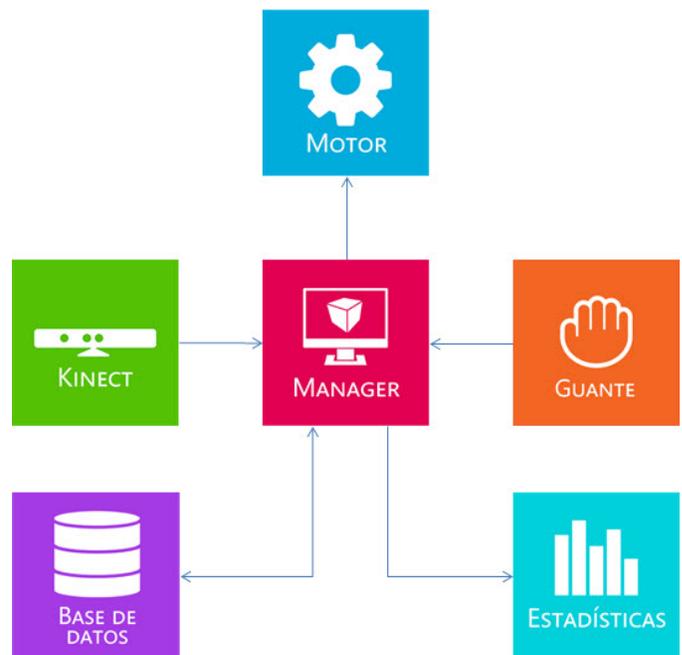


Figura 4: Representación de los Módulos que Conforman el Framework

1) **Controlador**: Este módulo constituye el núcleo del sistema, donde se encuentra las interfaces de programación definidas por el *framework*. Un videojuego se desarrolla bajo un API que permite la interacción y definición para todas sus funciones. Al mismo tiempo, el Controlador se encarga de la interacción de los otros módulos cada vez que sucede un evento que cambie entre los estados activo, pausado y cancelado.

2) *Motor*: El módulo de Motor agiliza y facilita la creación de los videojuegos o actividades terapéuticas al controlar los procedimientos necesarios para el despliegue gráfico. Así, es posible agrupar sus funcionalidades en: carga de modelos 3D y escenas complejas, gestor de animaciones y de iluminación.

La carga de modelos 3D realiza las operaciones requeridas para la lectura desde un archivo y el despliegue de un modelo 3D poligonal. Del mismo modo, la carga de escenas complejas, permite obtener el contenido general de una escena compleja formada por diversos modelos 3D. Particularmente, para nuestra propuesta se desarrolló un formato de escena basado en el uso de XML (Extensible Markup Language).

La carga de modelos y escenas solo toman en cuenta modelos estáticos, es decir, que no varían a través del tiempo. El gestor de animaciones permite realizar la carga de modelos con animación, permitiendo la presencia de dinamismo en las escenas sin la necesidad de haber interacción directa con el usuario. En nuestra propuesta, el formato de animación MD5 es empleado debido a su extenso uso en videojuegos.

Por último, el gestor de iluminación permite establecer y calcular las interacciones entre las fuentes de luz presentes en la escena, y los modelos y escenario. Para ello, existen diversos tipos de iluminación posibles: direccional, puntual, spot, y basada en la técnica Deferred lighting.

Esta última técnica de iluminación se realiza en dos pasadas, calculando la interacción lumínica en una escena. En una primera pasada se almacenan en *buffers* algunos datos de la escena como información de posición de los vértices, vectores normales, e incluso datos relativos a materiales de la escena desde el punto de vista de la cámara. Gracias al uso del hardware gráfico actual, es posible almacenar los *buffers* empleando texturas que residen en memoria de la tarjeta gráfica. En la segunda pasada, empleando la información de los diversos *buffers* se calcula de la forma en que cada píxel (en espacio de pantalla) resulta afectado por la luz directa o indirecta.

La Figura 5 muestra un ejemplo de la composición de los diversos *buffers* para obtener el resultado de la técnica e iluminación Deferred lighting.

Todas las técnicas de iluminación antes mencionadas son aplicadas con el objetivo de obtener un escenario 3D realista y recrear un ambiente confortable para los pacientes.

3) *Guante*: Nuestro *framework* está diseñado para el uso de actividades (videojuegos) empleando un guante virtual de datos para la captura del movimiento de los dedos y muñeca para el tratamiento de la motricidad fina.

El guante desarrollado por Ojeda et al. [14] se considera para ser integrado de forma funcional y sencilla a nuestra propuesta dado su simple interfaz de comunicación. Así, este módulo realiza la captura y tratamiento de imágenes desde una cámara web, a fin de detectar los datos necesarios para la interpretación de la movilidad de la mano.

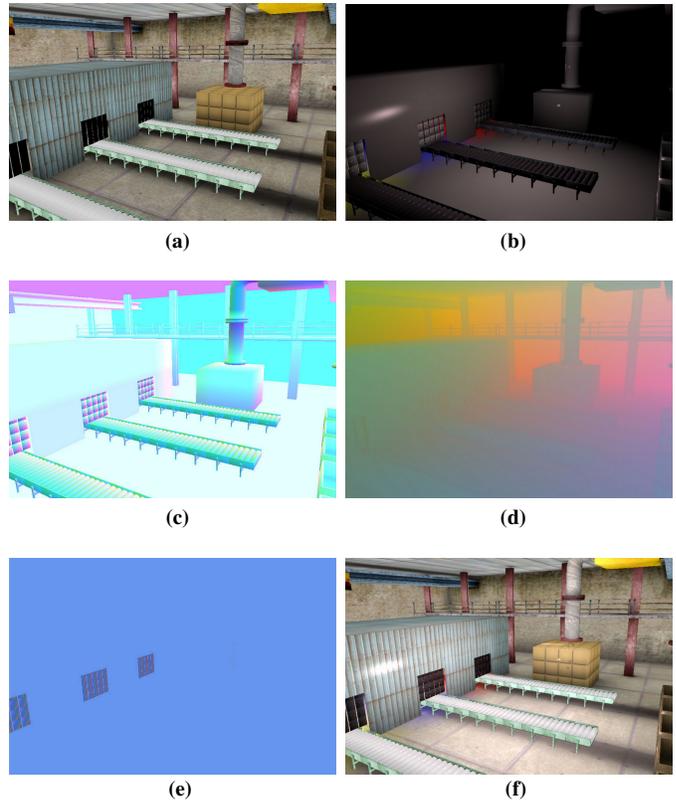


Figura 5: Técnica de Iluminación Deferred Lighting, donde se Muestran los *Buffers* de a) Valores Albedo, b) Mapa de Luces, c) Vectores Normales, d) Posición de los Vértices, e) Valores de Transparencia, y f) la Composición Final

El módulo extrae en tiempo real las imágenes desde una cámara web incorporada al guante para procesarlas empleando algoritmos de búsqueda basados en patrones cromáticos. Para ello se establecen de forma previa valores determinantes para esta operación como las dimensiones de la imagen a capturar, el color a rastrear dentro de cada imagen, y el tamaño mínimo de las áreas para ser consideradas como un patrón.

Una vez determinados los patrones cromáticos se extrae el valor de posición relativa de cada patrón en función a las dimensiones de la imagen, determinando el grado de flexión/extensión de cada dedo. Posteriormente, se hace correspondencia con cada uno de los dedos de la mano ejercitada (de izquierda a derecha en caso de emplearse la mano derecha, de derecha a izquierda en caso contrario). Finalmente, se reportan los valores de flexión/extensión detectados.

4) *Kinect*[®]: Este hardware de bajo costo se emplea como medio de captura de movimientos de gran extensión, procurando la ejercitación de la motricidad gruesa. Con este dispositivo se aprovecha la versatilidad de la realidad virtual no inmersiva en cuanto a la capacidad de aumentar la retroalimentación de acción-observación y así promover la activación cerebral a partir de las neuronas espejo.

Nuestra solución dispone de un conjunto de videojuegos desarrollados exclusivamente para el ejercicio de terapias de re-

habilitación, garantizando la especificidad de la ejercitación, la repetición y el entrenamiento cognitivo viso-espacial, aspectos que resultan fundamentales en métodos de re-entrenamiento.

Igualmente, se implementan metodologías de calibración de hardware a fin de ajustar el funcionamiento de los videojuegos del sistema a las capacidades específicas de cada paciente, permitiendo que la dinámica de la rehabilitación se ajuste a los rangos alcanzables de movilidad de cada paciente. Así, se despliegan gráficas analíticas con las cuales resulta posible establecer un seguimiento relativo a la evolución del paciente con respecto a la terapia. Además durante la ejecución de cada juego, se muestran gráficos que permiten al médico tratante evaluar el comportamiento del paciente.

5) *Estadísticas*: Para evaluar el estado y evolución del paciente en el transcurso del tiempo, se cuenta con un conjunto de metodologías que permite la generación automática de gráficos estadísticos: general, tiempo real e histórico.

Los gráficos generales constituyen estadísticas acumulativas con las que es posible evaluar el estado actual del paciente tomando en cuenta su actuación en el sistema. Los de tiempo real resultan de evaluaciones de la movilidad cuando realiza una sesión de rehabilitación empleando algunos de los videojuegos. Este tipo de gráficos se visualizan en tiempo de ejecución del videojuego. Por último, el histórico muestra de la actuación del paciente en cada sesión, permitiendo la evaluación de la evolución del mismo.

6) *Base de Datos*: El módulo manejador de base de datos es el encargado de gestionar todas las operaciones que requieren el acceso a la base de datos del sistema. Permite el registro, inicio de sesión y obtención de datos de un paciente previamente registrado, actualización y eliminación de un paciente. En nuestra propuesta, se emplea el manejador SQLite (<http://www.sqlite.org>).

V. PLATAFORMA DE REHABILITACIÓN

El *framework* desarrollado sobre nuestra propuesta de plataforma de rehabilitación ha sido implementado empleando el sistema operativo Windows 8, con el lenguaje de programación C#, el lenguaje de shading HLSL (High-Level Shader Language) y el IDE para desarrollo Microsoft Visual Studio 2012. Además se emplea el framework XNA en su versión 4.5. Adicionalmente se emplea un conjunto de herramientas de software como librerías y programas con los que se agilizan partes del desarrollo.

La población a la cual está dirigida nuestra plataforma son niños entre 6 y 12 años, debido al pensamiento lógico concreto alcanzado a dicha edad [15]. Debido a la variedad de patologías que acuden al Servicio de Rehabilitación Física del Hospital de Niños J.M. de Los Ríos se consideró clasificarlas en lesiones físicas-cognitivas y neurológicas. Se consideran pacientes que presentan limitaciones en amplitud articular, fuerza muscular, coordinación, déficit de atención, alteraciones sensorio-perceptuales y cognitivas.

A continuación se presenta cada una de las actividades

terapéuticas virtuales o videojuegos desarrollados en este trabajo.

A. Actividad Terapéutica La Fábrica

Este videojuego es el más complejo de los tres propuestos, mide, y estimula un mayor número de variables cognitivas. El videojuego se desarrolla en una fábrica, que dispone de tres bandas transportadoras. Cada banda transportadora está constituida por un conjunto de cilindros cuyas rotaciones genera una dinámica de transporte para los objetos sobre ésta.

Los objetos se deslizan sobre las bandas transportadoras, inicialmente situados en el extremo opuesto donde se encuentra el paciente con respecto a las bandas. Cada uno de estos objetos está asociado con uno de los seis modelos gráficos disponibles para el videojuego.

En el extremo cercano al jugador se dispone un conjunto de cajas que cumplen el papel de contenedores. El jugador tiene como objetivo recoger los objetos que se encuentran en las bandas transportadoras y ubicarlos en los contenedores. Los aciertos calificarán para la puntuación final. Cada contenedor, está identificado con el tipo de objeto que debe ser almacenado en él, información que debe tomar en cuenta el usuario antes de ubicar un objeto en los mismos. En la Figura 6 se puede apreciar la disposición de los elementos antes mencionados.



Figura 6: Captura de una vista del Videojuego La Fábrica dentro de nuestra Solución

Con La Fábrica se espera incrementar la capacidad de movimiento a lo largo y ancho del plano transversal, rompiendo un posible patrón de rigidez (o falta de fuerza) generado por atrofia del tono muscular. También se busca el movimiento de los miembros superiores actuando en contra de la gravedad. El videojuego proporciona al paciente una resistencia donde se pueden mejorar amplitudes, fuerza muscular y coordinación. Se determinó que el videojuego tiene un papel fundamental en el nivel cognitivo y perceptivo del jugador, donde el esquema corporal es necesario para poder precisar los movimientos que debe realizar con sus miembros superiores.

Este videojuego requiere de cuatro pasos para el cumplimiento de los objetivos generados. Específicamente la acción de recoger un objeto y colocarlo en el contenedor esperado es

generada mediante cuatro movimientos claves. La posición inicial del paciente consiste en permanecer en una posición bípeda con su miembro superior, derecho o izquierdo, en flexión de hombro horizontal y el otro miembro en flexión de hombro vertical.

Asumiendo que el brazo seleccionado (antes del inicio de la actividad) ha sido el brazo derecho, el jugador deberá:

- 1) Guiar la máquina de extracción hasta colocarla sobre un objeto que esté sobre alguna de las bandas transportadoras. Mover el brazo derecho paralelamente al plano transversal para colocar la máquina de extracción (cuya posición responderá a la posición del brazo) sobre la banda transportadora, procurando que existan objetos para ser extraídos. Así, se ejercita el hombro realizando movimientos de aducción y abducción horizontal.
- 2) Extraer el objeto de la cinta, levantando el brazo izquierdo con la finalidad de activar la máquina. Mantener el brazo izquierdo en esta postura hasta que el objeto sea capturado. Una vez el objeto es capturado, se proceder a relajar el brazo. Con este paso, se debe ejercer la extensión de hombros.
- 3) Colocar la máquina de extracción sobre el contenedor correspondiente. Una vez extraído el objeto, colocar la máquina sobre el contenedor correspondiente. Ejercitando nuevamente el hombro a partir de movimientos de aducción y abducción horizontal.
- 4) Dejar caer el objeto en el contenedor al estar la máquina colocada sobre el contenedor correspondiente al objeto capturado, y mantener el brazo extendido con una flexión de hombro de cero grados. Luego ejercer un movimiento de flexión de codo y posterior a éste regresar al estado inicial.

B. Actividad Terapéutica GALAXY

La temática del videojuego GALAXY sitúa al jugador en una nave espacial que está siendo embestida por un conjunto de asteroides. En la Figura 7 se pueden apreciar todos los elementos que componen el videojuego.

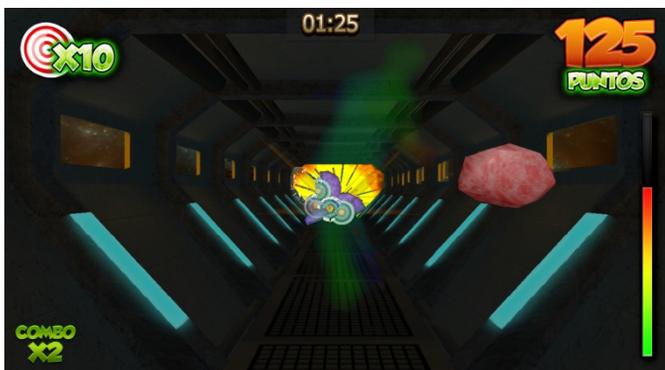


Figura 7: Captura de un Instante del Videojuego GALAXY en su Ejecución

Se asume que la nave ha perdido una de sus partes debido al impacto de un asteroide. El jugador tiene la misión de evitar que nuevos asteroides ingresen a la nave re-dirigiéndolos fuera

de ella. A medida que los asteroides impactan la nave, trozos de éstos podrían entrar en ella. El jugador deberá usar sus manos, para guiarlos fuera de la misma.

Además de cumplir la tarea de expulsar los asteroides que tiene una puntuación en el juego, el jugador tiene la oportunidad de mejorar su desempeño impactando nuevos objetivos en forma de dianas que aparecerán periódicamente en escena. En esta actividad terapéutica se trabajan ambos brazos al mismo tiempo, ejercitando la motricidad gruesa, así como la coordinación al lograr alcanzar el asteroide y redirigirlo en la dirección deseada. El paciente deberá extender ambos brazos en paralelo al plano transversal y frontal, en su máxima capacidad posible, para poder abarcar todas las áreas en las que los asteroides pueden aparecer. Logrando el paciente realizar ejercicios de extensión y coordinación viso-espacial. Debido a que el paciente tendrá la necesidad de realizar movimientos rápidos para alcanzar a golpear todos los asteroides en la escena, se desea que el mismo realice ejercicios de fuerza durante el transcurso de la actividad.

En GALAXY se hace especial énfasis en ejercitar el movimiento del brazo a lo largo del plano sagital (flexión-extensión del codo), ya que el paciente obtendrá un mayor puntaje dependiendo de la fuerza con la que golpee el asteroide, la cual es determinada por la velocidad con la que el jugador pueda extender el brazo en el eje Z hasta la máxima extensión registrada en el calibrador. No existen gestos predefinidos para jugar, el jugador puede mover sus manos libremente para alcanzar los objetivos.

El videojuego está diseñado para ejercitar la flexión-extensión del codo mediante el movimiento que debe realizar el paciente para poder aplicar fuerza a los asteroides, los movimientos que implican alcanzar los asteroides para impactarlos ejercitan la amplitud articular del brazo.

Debido a su característica de no poseer limitantes en los movimientos y el desplazamiento libre, el videojuego está indicado para ser utilizado en pacientes con cualquier tipo de patología que involucre problemas motores dentro de la terapia de rehabilitación.

Como el paciente debe tener la capacidad de coordinar sus movimientos con las acciones que ocurren en la escena y requiere de un cierto grado de concentración, el juego sirve a su vez como un instrumento para tratar a pacientes con desordenes de atención y autismo. Se logra un movimiento bilateral en los miembros superiores, fortalecimiento de la fuerza muscular por la posición que se debe mantener para la ejecución del juego donde actúan en contra de la gravedad.

El videojuego proporciona al paciente una movilidad en sus miembros superiores, motricidad gruesa por el desplazamiento mientras ejecuta el juego. Logrando que realice una integración bilateral inhibiendo patrones posturales inadecuados y facilitando la funcionalidad de ambos miembros.

Este juego proporciona un nivel de atención y concentración para mejorar la capacidad cognitiva del paciente.

C. Actividad Terapéutica ACIS

La temática de este videojuego se desarrolla en un ambiente acuático con elementos tales como, peces y un pescador. El jugador debe capturar con una caña de pescar los peces que viajan de un lado a otro en la escena. En ACIS, la cámara de visualización 3D se dispone en una posición tal que, en compañía a los elementos que conforman la escena, simula estar bajo el agua. Un ejemplo de ello se muestra en la Figura 8.



Figura 8: Instante dentro del Videojuego ACIS Ejecutado por un Jugador

Los distintos peces son representados con modelos gráficos. Para ello, se define una estructura de datos genérica para la representación de un pez que tiene asociado un comportamiento y una representación gráfica, única para cada tipo de pez. Igualmente, la estructura almacena el peso del pez.

El jugador controla un anzuelo que es el elemento con cual podrá capturar los peces. Éste está centrado en la escena y se desplaza sobre el eje de ordenadas. La captura de un pez es valuada con una puntuación que varía según el peso del mismo, el cual será percibido por el jugador por el tiempo que tarda en sacar virtualmente el pez del agua.

Con este juego se busca ejercitar el hombro a partir de la flexión y extensión del mismo. Se desea motivar al paciente a realizar tal movimiento en su máxima extensión, alentándolo incluso a que supere sus límites de movilidad (que fueron calibrados previamente en la etapa de calibración del Kinect®).

La interacción del jugador con ACIS está conformada en dos fases. La primera consiste en posicionar el anzuelo en la trayectoria del pez, esto se logra con movimientos sobre el eje de ordenadas. En la segunda fase el jugador debe sacar el pez del agua, para ello deberá ejercer una flexión de hombro prolongada hasta lograr el objetivo, el tiempo de esta flexión dependerá del peso virtual del pez.

Este videojuego sólo involucra un tipo de movimiento físico para el paciente, procurando de esta manera disminuir la carga cognitiva. De esta forma, resulta un videojuego simple y efectivo desde el punto de vista terapéutico.

VI. EVALUACIÓN DEL SISTEMA

Durante la evaluación de un sistema computacional, se realizan diversas pruebas para comprobar su efectividad. Una de las

pruebas consiste en la evaluación del diseño de la interfaz de usuario. Para ello, Nielsen [16] propone una serie de heurísticas tales como:

- A Visibilidad del estado del sistema
- B Utilización del lenguaje de usuario
- C Control y libertad para el usuario
- D Consistencia y estándares
- E Prevención de errores
- F Minimización de la carga cognitiva
- G Flexibilidad y eficiencia de uso
- H Presencia de diálogos estéticos y diseño minimalista
- I Ayuda a los usuarios en el reconocimiento, diagnóstico y recuperación ante errores
- J Ayuda y documentación

Con el objetivo de demostrar la capacidad de nuestra solución, se aplicaron estas heurísticas en forma de encuesta. En dicha encuesta se pregunta a un total de 5 personas, cada uno de los tópicos (de la A-J) presentados de forma explicativa, donde la respuesta posible es SI o NO. La Figura 9 muestra los resultados luego de aplicar las encuestas.

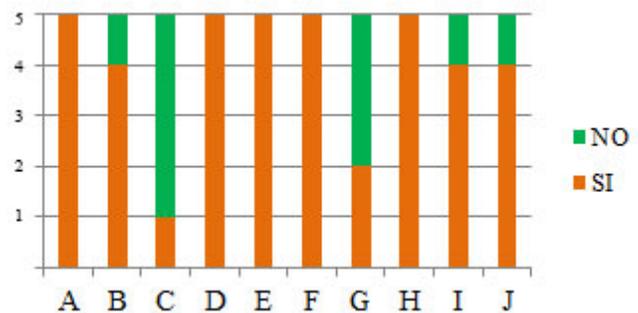


Figura 9: Resultado de la Encuesta Aplicada sobre la Heurística de Nielsen

Los datos obtenidos reflejan la alta conformidad de los usuarios con el sistema. El punto C es de carácter restrictivo con respecto al libre manejo de parte del usuario por estar en un ambiente supervisado. El comportamiento del punto G es debido a la necesidad de la prosecución de los procesos en el sistema. Cabe destacar que el sistema se encuentra en mejoramiento continuo.

VII. DISCUSIÓN Y CONCLUSIONES

En este trabajo se presenta una arquitectura robusta para la rehabilitación física en miembros superiores empleando realidad virtual. Para ello, se presenta una estructura de hardware y software soportado en el desarrollo de un *framework* extensible y simple que maneja la creación de diversos videojuegos dentro de nuestra arquitectura.

Así, se presenta un sistema de realidad virtual de bajo costo para uso clínico, en pacientes entre 6 y 12 años con discapacidades en miembros superiores, del sistema hospitalario nacional. Siendo una solución reutilizable para miembros inferiores u otras discapacidades de patologías particulares. Al mismo tiempo, el uso del sistema no representa ningún riesgo para el paciente que se somete a éste.

El sistema utiliza una plataforma de hardware sencilla y de fácil construcción, haciéndolo escalable, reutilizable y robusto. Es importante destacar que el sistema no requiere de entrenamientos especiales o costosos por parte del personal médico, siendo controlado en todo momento por estos pudiendo intervenir en cualquier momento de su ejecución.

Se realizaron sesiones de prueba con personal médico especializado teniendo una gran aceptación por parte del mismo. Sin embargo, se observó el cansancio del jugador al permanecer de pie por un largo tiempo. Actualmente, se tiene proyectado el uso de nuestra solución en pacientes con discapacidad móvil (leve o severa) en sus miembros inferiores. Sin embargo, se han realizado pruebas con usuarios sentados.

El trabajo aquí presentado constituye la primera fase de un proyecto interdisciplinario desarrollado para el Hospital de Niños J.M. de Los Ríos en Caracas, Venezuela. Como segunda fase está la implantación del sistema en las instalaciones del hospital e incorporar videojuegos que incluyan la motricidad fina con un guante de datos. En una tercera fase se proyecta evaluación del sistema en pacientes y estudiar su progreso terapéutico en el tiempo con el uso del sistema.

AGRADECIMIENTOS

Al personal del Servicio de Medicina Física y Rehabilitación del Hospital de Niños J.M. de Los Ríos. A los niños y niñas, pacientes y sus padres por habernos permitido obtener información valiosa para la realización de este trabajo. Igualmente, a Daniellys Bolívar de 6 años quien ha jugado repetidas veces los videojuegos presentados en este trabajo.

REFERENCIAS

- [1] Organización Mundial de la Salud y Banco Mundial, *Informe Mundial sobre la Discapacidad*, Ediciones de la OMS, 2011. <http://tinyurl.com/bt4xes9>

- [2] M. Zepeda y A. Vásquez, *Aplicación de la Clasificación Internacional del Funcionamiento, de la Discapacidad y de la Salud en Estudios de Prevalencia de Discapacidad en las Américas*, Organización Panamericana de la Salud, 2012.
- [3] G. Burdea y P. Coiffet, *Tecnologías de la Realidad Virtual*, Barcelona, España: Paidós, 1996.
- [4] H. Sveistrup, *Motor Rehabilitation using Virtual Reality*, Journal of NeuroEngineering and Rehabilitation, vol. 1, no. 1, pp. 10, 2004.
- [5] J. Broeren, H. Samuelsson, K. Stibrant-Sunnerhagen, C. Blomstrand y M. Rydmark, *Neglect Assessment as an Application of Virtual Reality*, Acta Neurologica Scandinavica, vol. 16, no. 3, pp. 157-163, Septiembre 2007.
- [6] M. Holden, *Virtual Environments for Motor Rehabilitation: Review*, CyberPsychology & Behavior, vol. 8, no. 3, pp. 187-211, 2005.
- [7] J. Webster, P. McFarland, L. Rapport, B. Morrill, L. Roades y P. Abadee, *Computer-Assisted Training for Improving Wheelchair Mobility in Unilateral Neglect Patients*, Archives of Physical Medicine and Rehabilitation, vol. 82, no. 6, pp. 769-775, Junio 2001.
- [8] A. Merians, H. Poizner, R. Bolan, G. Burdea y S. Adamovich, *Sensorimotor Training in a Virtual Reality Environment: Does it Improve Functional Recovery Poststroke?*, Neurorehabilitation and Neural Repair, vol. 20, no. 2, pp. 252-267, Junio 2006.
- [9] S. You, S. Jang, Y.-H. Kim, Y.-H. Kwon, I. Barrow y M. Hallettm, *Cortical Reorganization Induced by Virtual Reality Therapy in a Child with Hemiparetic Cerebral Palsy*, Developmental Medicine & Child Neurology, vol. 47, no. 1, pp. 628-635, 2005.
- [10] Bienetec, *Biotrak*, <http://www.biotraksuite.com>
- [11] Universitat Pompeu Fabra, *Rehabilitation Gaming System*, España. <http://rgs-project.eu>
- [12] C. Norman, D. Clark y C. Barnaby, *Kinesthesia*, National Instruments. <http://tinyurl.com/7f67ts6>
- [13] K. LaBelle, *Evaluation of Kinect Joint Tracking for Clinical and In-Home Stroke Rehabilitation Tools*, Notre Dame Thesis, 2011.
- [14] J. Ojeda, E. Ramírez, C. Mena y O. Rodríguez, *A Low-Cost Data Glove for Virtual Reality* in proceedings of the XII International Congress of Numerical Methods in Engineering and Applied Sciences (CIMENICS), pp. 19-24, 2014.
- [15] H. Ginsburg y S. Opper, *Piaget's Theory of Intellectual Development*, Prentice Hall, 1979.
- [16] J. Nielsen, *Enhancing the Explanatory Power of Usability Heuristics*, in proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 152-158, 1994.

Revisión Documental Sistemática de Enfoques y Técnicas para la Construcción de Arquitecturas en un Contexto de Líneas de Productos de Software

Juan Herrera¹, Alfredo Matteo², Francisca Losavio²
jchr1982@gmail.com, alfredojose.matteo@gmail.com, francislosavio@gmail.com

¹ PFG Informática para la Gestión Social, Universidad Bolivariana de Venezuela, Caracas, Venezuela

² Escuela de Computación, Universidad Central de Venezuela, Caracas, Venezuela

Resumen: Las líneas de productos de software son un enfoque centrado en el desarrollo de software basado en la reutilización, mediante el proceso de ingeniería de líneas de productos y particularmente en la fase de ingeniería del dominio es donde se diseña la arquitectura de referencia y/o la arquitectura de líneas de productos, para ello dos enfoques son utilizados en la industria, partiendo de un amplio conocimiento del dominio de la aplicación se encuentra el enfoque proactivo y partiendo de los sistemas existentes se encuentra el enfoque reactivo, es de particular interés identificar los artefactos utilizados en las estrategias utilizadas por ambos enfoques diferenciándolos ya sea por el tipo de reutilización seleccionado o por la incorporación de nuevos artefactos en cada uno de los métodos propuestos, en consecuencia una revisión documental sistemática de los estudios que abordan el tema es indispensable para determinar los alcances y limitaciones de las propuestas encontradas.

Palabras Clave: Líneas de Productos de Software; Arquitectura de Referencia; Arquitectura de Líneas de Productos; Proactivo; Reactivo; Revisión Documental Sistemática.

Abstract: The software product lines are a software development approach based on reuse, through the process of product lines engineering; particularly in the domain engineering phase is where the reference architecture and/or product line architecture is designed. For this purpose, two approaches are commonly used in industry, one based on extensive knowledge of the application domain or proactive approach, and the other one, or reactive approach, based on building on existing systems; in particular, it is interesting to identify the artifacts used in the strategies employed by both approaches, differentiating either for the selected type of reuse or incorporation of new artifacts in each of the proposed methods. In consequence, a systematic literature review of studies that address this issue is essential to determine the scope and limitations of the proposals found.

Keywords: Software Product Line; Reference Architecture; Product Line Architecture; Proactive; Reactive; Systematic Literature Review.

I. INTRODUCCIÓN

El enfoque denominado LPS (Líneas de Productos de Software) se centra en el desarrollo de software basado en la reutilización, que se fundamenta en la identificación y captura de las similitudes y variabilidades de productos de software dentro de un dominio determinado [1]; con ello se pretende reducir el tiempo, esfuerzo, costo, complejidad y una mejora de la productividad en el desarrollo y mantenimiento de sistemas de software complejos [2].

Es a través de la ILPS (Ingeniería de Líneas de Productos de Software) donde la reutilización es planificada y aplicada [1]. La ILPS ha demostrado ser un paradigma para el desarrollo de una diversidad de sistemas intensivos de software utilizando

tecnologías existentes para la construcción a gran escala de productos adaptados a las necesidades individuales de los clientes a un menor costo, en menor tiempo y con mayor calidad [1].

El paradigma de ILPS comprende dos grandes procesos [1]: ID (Ingeniería del dominio) que es responsable de establecer la plataforma reutilizable y de definir las similitudes y variabilidades de la línea de productos. IA (Ingeniería de la aplicación) que se encarga de derivar aplicaciones de líneas de productos a partir de la plataforma establecida en la ingeniería de dominio.

Las actividades incluidas en el proceso de ID son la gestión del producto, la ingeniería de requisitos del dominio, el diseño del

dominio, la realización del dominio y las pruebas del dominio [3]. La gestión del producto se ocupa de los aspectos económicos, en particular de las estrategias de mercado relacionadas con las líneas de productos. La ingeniería de requisitos del dominio es responsable de la identificación de los requisitos comunes y variables. El diseño del dominio define la arquitectura de referencia de la línea de productos proporcionando una estructura común de alto nivel para todas las aplicaciones derivadas. La realización del dominio se ocupa del diseño detallado y la implementación de componentes de software reutilizables. Las pruebas del dominio revelan la evidencia de los defectos en los artefactos del dominio y crea artefactos de prueba reutilizables para la evaluación de las aplicaciones. De allí, para el propósito de esta investigación, solo es de interés la actividad relacionada con el diseño del dominio y particularmente con el artefacto que se deriva de esta actividad que es la *arquitectura del dominio*. El *dominio* es un campo de estudio que define un conjunto de requisitos comunes, la terminología y la funcionalidad de cualquier programa de software construido para resolver un problema.

En este contexto, se pueden encontrar dos tipos de arquitecturas de software para el dominio: AR (Arquitectura de Referencia) y ALP (Arquitectura de Líneas de Productos). Ambas apuntan a mejorar el desarrollo del software mediante la estandarización de las arquitecturas de un conjunto de sistemas de software [4].

En general, una *arquitectura de referencia* acumula conocimiento de un dominio, identifica soluciones abstractas de un problema y promueve la reutilización de las experiencias en diseño alcanzando una sólida y bien reconocida comprensión y conocimiento de un dominio, ofreciendo soluciones estandarizadas para un dominio de aplicación más amplio [4][5]. Es generalmente construida siguiendo un enfoque denominado top-down, descendente o proactivo [6].

Teniendo en cuenta la importancia de las ARs, diferentes dominios de aplicaciones, como el automotriz, la aeronáutica y la robótica, han propuesto y utilizado estas arquitecturas [4].

La *arquitectura de líneas de productos*, en cambio, es más especializada, centrándose en un subconjunto específico de los sistemas de software existentes de un dominio y ofrecen soluciones estandarizadas para una familia pequeña de sistemas [4], lo que permite crear sistemas de forma más confiable y rentable [2].

Una diferencia esencial es que las ALPs tratan con la variabilidad entre los productos, y se sitúan en un nivel de abstracción más bajo en comparación con las ARs [4].

La ALP es también una arquitectura genérica, pero es construida a partir de productos existentes, por lo tanto “hereda” de cierta forma las propiedades arquitectónicas, por ejemplo el estilo arquitectónico, y las funcionalidades relevantes de cada producto considerado. Por eso es construida mediante un enfoque denominado bottom-up, ascendente, extractivo o reactivo [6][7], que considera técnicas de reingeniería en diferentes etapas del desarrollo. Es considerado un enfoque útil en la práctica, cuando en la industria no se tiene una LPS y se quiere construir una, a partir de productos que funcionan y han sido elaborados anteriormente; es a partir de ellos que se quiere generalizar y construir la ALP. Las ARs

están generalmente situadas en un nivel de abstracción más alto en comparación con las ALPs [4].

A partir de aquí, el análisis y alcance del desarrollo de la AR y ALP en esta investigación se caracteriza por identificar los métodos y los artefactos que se utilizan y/o reutilizan para construir la arquitectura (AR o ALP) a partir de dos tipos de enfoques: proactivo y reactivo [7], su justificación se basa en las nuevas tendencias en el desarrollo de las líneas de productos como medio para la personalización masiva del software y de allí los dos enfoques para la construcción de las arquitecturas, estos modelos son: el enfoque proactivo, y el enfoque reactivo [6][7].

En consecuencia el interés de este trabajo consiste en considerar la AR o ALP, centrándose en los enfoques proactivo y reactivo.

El objetivo principal de este trabajo es explorar, organizar, resumir y analizar las principales tendencias que se han propuesto o utilizado en la construcción de las ARs y/o ALPs. Para ello, se ha adoptado y llevado a cabo una revisión sistemática bibliográfica, mediante la guía proporcionada por Kitchenham para estudiar y analizar la literatura en el campo de la ingeniería del software [8][9][10].

Este trabajo está estructurado de la siguiente manera, además de esta introducción y las conclusiones: en la Sección II se presentan los enfoques y modelos a considerar; en la Sección III se presenta el análisis de la revisión documental sistemática realizada. En la Sección IV se discuten los resultados obtenidos. Finalmente, en la Sección V se da un resumen de los aportes de la investigación.

II. ENFOQUES Y MODELOS A CONSIDERAR

A. Personalización Masiva de Software y sus Beneficios

La *personalización masiva* del software es un enfoque que se concentra en los medios de producción eficiente y el mantenimiento de múltiples productos de software similares, aprovechando lo que tienen en común y gestionando las variantes entre ellos. El objetivo es que el enfoque cambie del desarrollo y mantenimiento de varios productos distintos, hacia el desarrollo y mantenimiento de una sola LPS [7]. Cabe decir, que la personalización masiva es la producción a gran escala de productos adaptados a las necesidades individuales de los clientes [1].

B. Enfoques para la Adopción de la Personalización Masiva de Software

La transición de las organizaciones a la personalización masiva de software puede lograrse con uno de los tres enfoques de amplia adopción: proactivo, reactivo, y extractivo [7].

El *enfoque proactivo* es apropiado cuando los requisitos para el conjunto de los productos necesarios y futuros, están bien definidos y estables. Este enfoque requiere un esfuerzo considerablemente para la organización, pero esto cae bruscamente una vez que la línea de producción se ha completado [1][7]. Las tareas de alto nivel que se llevan a cabo son: a) realizar un análisis de dominio y del alcance para identificar la variación que será soportada en la línea de producción; b) modelar la arquitectura de la línea de productos para soportar todos los productos en la línea de producción y c) diseñar las partes comunes y variables del sistema [7].

En el *enfoque reactivo*, la organización aumenta su línea de productos debido a la demanda de nuevos productos o cuando surgen nuevos requisitos a los productos existentes. Las tareas de alto nivel que se llevan a cabo son: a) caracterizar los requisitos para el nuevo producto en relación con los actualmente soportados en la línea de producción y b) si el nuevo producto está dentro del alcance de la línea de producción actual, se crea la definición del producto, en caso contrario, se amplía el alcance de la línea de producción actual para incluir los nuevos requisitos [7].

En el *enfoque extractivo*, la organización aprovecha los sistemas existentes de software mediante la extracción de lo común y lo diverso del código fuente, para sintetizarlos en una sola línea de producción. Requiere utilizar técnicas de reingeniería para reconstruir la arquitectura del sistema a partir del código fuente, o de la documentación existente. Las tareas de alto nivel que se llevan a cabo son: a) identificar elementos comunes y la variación en los sistemas existentes y b) factorizar en una sola línea de productos [7].

En esta investigación solo se utilizarán los enfoques proactivo y reactivo. El enfoque extractivo es considerado como similar al enfoque reactivo, ya que en ambos se reutilizan los sistemas existentes [1]: sistemas de software ad hoc y/o LPS presentes en la organización.

Las tareas que van a servir de base en el proceso de la toma de decisiones para determinar en cada uno de los estudios si el método de construcción de la arquitectura utiliza el enfoque proactivo (similar al enfoque top-down, según el proceso de desarrollo de software tradicional [11]) o por el contrario, utiliza el enfoque reactivo (similar al enfoque bottom-up [11]) son:

Se considera que el método pertenece al enfoque proactivo si describe las siguientes tareas en este orden: 1) análisis del dominio, 2) definición del alcance determinado por la variedad de productos que la organización considera que van pertenecer a ese dominio y 3) el modelado de la arquitectura.

Se considera que el método pertenece al enfoque reactivo si describe las siguientes tareas en este orden: 1) elige el conjunto de sistemas existentes que formarán parte de la línea de productos, 2) identifica las similitudes y diferencias entre los sistemas existentes y 3) el modelado de la arquitectura.

III. REVISIÓN DOCUMENTAL SISTEMÁTICA

La revisión sistemática se realizó en el contexto de la AR y ALP, con el objetivo de identificar todos los estudios relevantes relacionados con los métodos de construcción de estas arquitecturas. Se llevó a cabo en el primer trimestre del 2013. Kitchenham indica que la gestión del proceso de revisión sistemática se fundamenta en tres fases principales: Planificación de la revisión, Realización de la revisión e Informe de la revisión [8][9][10].

En esta revisión documental sistemática se pretende conocer y comprender los enfoques y/o técnicas de diseño de las ARs y ALPs, particularmente el interés se centra en si aplican el enfoque proactivo o reactivo en la construcción de estas arquitecturas, además, de determinar el nivel de reutilización con relación a métodos y artefactos arquitecturales tradicionales o por el contrario si los métodos describen nuevos métodos y/o artefactos para la construcción de estas. Por

último identificar las similitudes y diferencias entre las ARs y ALPs.

En la Tabla I se describen las fases, las etapas que incluyen cada fase y los artefactos generados en el proceso de revisión sistemática.

Tabla I: Proceso de Revisión Documental Sistemática

Fases	Etapas	Artefactos
Planificación de la revisión	Justificación de la necesidad de una revisión.	Protocolo
	Especificación de la(s) interrogante(s) de investigación.	
	Desarrollo del protocolo de revisión.	
Realización de la revisión	Identificación de los estudios relacionados.	Estudios aceptados
	Selección de los estudios primarios.	
	Evaluación de la calidad de los estudios.	
	Extracción de los datos.	Modelo de datos
	Síntesis de los datos.	
Informe de la revisión	Especificación de los mecanismos de difusión.	Tablas resumen
	Presentación de los resultados	

A. Planificación de la Revisión

Esta fase define los objetivos de la investigación y el protocolo en que la revisión se ejecutará.

1) *Identificación de la Necesidad para una Revisión Sistemática:* Como parte de un proyecto de investigación mayor que investiga la identificación y caracterización de las ARs y ALPs debido a que ambas arquitecturas tienen definidos sus roles en lo que respecta a objetivos específicos de las arquitecturas de software y en particular para el desarrollo de las LPS. La identificación y comprensión de enfoques y/o técnicas existentes que apoyan la construcción de estas arquitecturas es relevante para comprender sus alcances y limitaciones.

2) *Desarrollo del Protocolo:* El protocolo aborda las preguntas de investigación, las estrategias de búsqueda, criterios y procesos de selección, los criterios de evaluación de la calidad, el modelo y proceso de extracción de datos, y el plan de análisis de los datos. En otras palabras, el protocolo es un plan o conjunto de pasos a seguir en un estudio.

Las preguntas de investigación tienen el objetivo de encontrar todos los estudios para comprender y resumir las evidencias acerca de los enfoques propuestos o utilizados para construir las ARs y ALPs, las siguientes PIs (Preguntas de Investigación) son consideradas:

- PI1. ¿Cuáles son los métodos utilizados para el diseño de las ARs y ALPs tomando en consideración el enfoque proactivo (top-down) o reactivo (bottom-up) como estrategias para el diseño de las arquitecturas?
- PI2. ¿Los métodos reutilizan o adaptan algún patrón, estilo o modelo arquitectónico tradicional o describen nuevos métodos o artefactos para la construcción de las arquitecturas?

- PI3. ¿Cuáles son las diferencias encontradas entre las ARs y las ALPs?

Adicionalmente, otras preguntas fueron formuladas:

- PI2.1. ¿Cuál es la forma o estrategia de reutilización más utilizada?
- PI2.2. ¿Cuál de los dos modelos (proactivo y reactivo) es el más utilizado?
- PI2.3. ¿Cuáles otros aspectos fueron evidenciados en los métodos analizados?

Con relación a la estrategia de búsqueda, primero se establecieron las fuentes de búsqueda para encontrar los estudios primarios, particularmente se seleccionó el motor de búsqueda ResearchGate¹ como la única fuente de búsqueda por ser una herramienta de investigación científica, que permite a los investigadores buscar contenido en revistas arbitradas, conferencias, workshop y repositorios institucionales con información relevante para la realización de la revisión sistemática.

Para establecer la estrategia de búsqueda, teniendo en cuenta las preguntas de investigación, inicialmente se identificaron las principales palabras clave “Reference Architecture” y “Product Line Architecture”. En la Tabla II se encuentran todas las expresiones de búsqueda utilizadas en la revisión sistemática.

Tabla II: Expresiones de Búsqueda en la Revisión Sistemática

Fuente consultada = “http://www.researchgate.net”		
Expresiones de búsqueda	Hits	Aceptados
(“product line architecture” OR “product line architectures”) AND (“approach”)	91	29
(“product line architecture” OR “product line architectures”) AND (“method” OR “methods”)	48	
(“product line architecture” OR “product line architectures”) AND (“proactive”)	5	
(“product line architecture” OR “product line architectures”) AND “reactive”	2	
(“product line architecture” OR “product line architectures”) AND “top-down”	0	
(“product line architecture” OR “product line architectures”) AND “bottom-up”	0	
Total	146	
(“reference architecture” OR “reference architectures”) AND (“approach”)	210	40
(“reference architecture” OR “reference architectures”) AND (“method” OR “methods”)	100	
(“reference architecture” OR “reference architectures”) AND “proactive”	1	
(“reference architecture” OR “reference architectures”) AND “reactive”	2	
(“reference architecture” OR “reference architectures”) AND “top-down”	3	
(“reference architecture” OR “reference architectures”) AND “bottom-up”	3	
Total	319	
Total general	465	69

Con relación a los criterios de inclusión y exclusión. Criterios por los cuales los estudios serán evaluados para decidir si deben ser seleccionados o no en el contexto de la revisión sistemática.

Por lo tanto, los CIs (Criterios de Inclusión) que se utilizan para incluir los estudios relevantes en esta revisión sistemática fueron:

- CI1: El estudio define AR y/o ALP en el ámbito de las LPS.
- CI2: El estudio propone o emplea un método para construir las ARs y/o ALPs en el ámbito de las LPSs.

Los CE (Criterios de Exclusión) que se utilizan para excluir los estudios que no contribuyen a responder a las interrogantes de la investigación son:

- CE1: El estudio no propone o utiliza el modelo proactivo (top-down) y/o reactivo (bottom-up) para construir las ARs y/o ALPs;
- CE2: El estudio donde su texto completo no esté disponible o solo aparece el resumen; y
- CE3: El estudio que esté escrito en un idioma diferente al inglés o español.

Con respecto a la extracción de datos y método de síntesis. Se prevé la construcción de tablas para la obtención de datos relacionados con las preguntas de investigación, en caso de que sea aplicable.

En la selección inicial de las publicaciones de entre los resultados de búsqueda, se dio lectura a los títulos y resúmenes de las publicaciones encontradas. En la Figura 1 se muestra el modelo de datos inicial utilizado en la revisión sistemática. Las publicaciones fueron almacenadas en carpetas independientes de acuerdo a la expresión de búsqueda utilizada, de esta forma facilitar la ubicación de los estudios para realizar el análisis con mayor fluidez y comodidad.

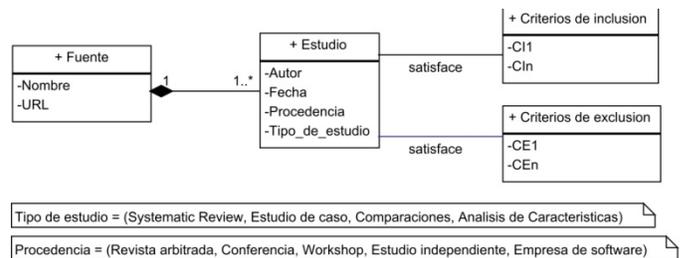


Figura 1: Modelo de Datos Inicial para la Revisión Sistemática

De las 69 publicaciones definitivamente seleccionadas se procede a la extracción de la información relevante utilizando el modelo de datos que será utilizado en las fases de realización e informe de la revisión representado en la Figura 2, y que incluye la lectura de sus contenidos. En la recolección de datos, las citas textuales de los artículos analizados se conservarán en el idioma original (inglés) para evitar el posible sesgo al traducir y posteriormente parafrasear según la interpretación del investigador sobre la cita leída. Esto permite una vez realizados los aportes a la investigación, poder referirse a las citas por posibles malas interpretaciones de las mismas, para validar y para futuras referencias. No obstante, al final de la revisión sistemática los resultados finales reflejados en la fase de informe de la revisión se colocarán en español que es el idioma nativo del investigador.

¹ http://www.researchgate.net

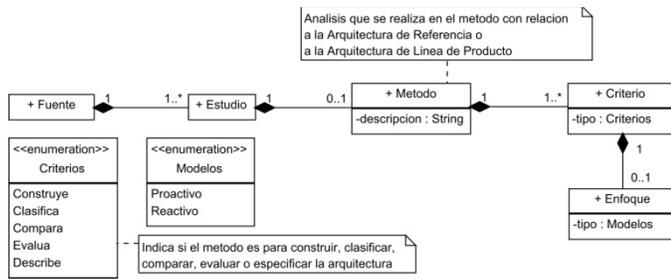


Figura 2: Modelo de Datos para el Vaciado de los Datos

Para estandarizar la forma en que la información estará representada, se definió un formato para recopilar datos de los estudios seleccionados. La Tabla III muestra el formato que será utilizado para registrar los resultados de la extracción de los datos, tomando en cuenta el modelo de datos de la Figura 2.

Tabla III: Formato para la Recolección de Datos

Criterio	Estudio		Método	TA	Asuntos de interés en relación a los artefactos utilizados por el método		
	Año	Ref			Reutiliza y adapta	Reutiliza tal cual	Agrega artefacto

La Tabla IV describe las propiedades de extracción que serán utilizados a lo largo de la revisión sistemática.

Tabla IV: Propiedades de Extracción

Propiedad	Valor	Descripción
Criterios	Construye	Construyen la arquitectura bajo el enfoque reactivo (bottom-up); o proactivo (top-down).
	Clasifica	Clasifican las arquitecturas
	Compara	Comparan métodos de construcción arquitectural.
	Evalúa	Evalúan la calidad arquitectural.
	Especifica	Describen formalmente la arquitectura.
Estudios	Año	Año de publicación del estudio.
	Ref.	Referencia bibliográfica y/o webgráfica.
Método	Nombre	Nombre del método identificado en el estudio.
TA	Tipo de arquitectura	Indica si el método se basa en la AR o ALP.
Asuntos de interés en relación a los artefactos utilizados por el método	Reutiliza y adapta	Indica si el método, técnica, patrón, estilo, estándar o framework es reutilizado y adaptado para construir la arquitectura.
	Reutiliza tal cual	En este caso los artefactos son reutilizados sin modificarlos para construir la arquitectura.
	Agrega artefacto	En este caso se crea o propone algún nuevo método, técnica, patrón, estilo, estándar o framework para construir la arquitectura.

B. Realización de la Revisión

En esta fase, los estudios son identificados, seleccionados y evaluados de acuerdo con el protocolo previamente establecido. Como resultado de la búsqueda en las base de datos de publicaciones con las expresiones de búsquedas, 465 estudios fueron localizados, de los cuales solo 69 estudios fueron aceptados para su análisis.

Para la selección final de los estudios aceptados se realizó un proceso filtrado que consistió en la aplicación de los criterios de inclusión y de exclusión, en la Tabla V se muestra un resumen de los resultados indicando la cantidad en números y porcentajes de los estudios y que a manera de resumen fueron

agrupados en los dos elementos fundamentales comunes en los criterios de búsqueda expresados en la Tabla II.

Tabla V: Relación de Estudios por Criterios de Búsqueda

Criterios de búsqueda	Artículos			
	Incluidos	Excluidos	Total de aciertos	% Incluidos
Reference Architecture	40	279	319	57.97
Product Line Architecture	29	117	146	42.03
Total	69	396	465	

La Tabla VI presenta la relación de los estudios cuantificados por año y tipo de publicación, de los 69 estudios que se consideraron pertinentes para esta revisión documental sistemática.

Tabla VI: Relación de Estudios por Años y Tipos de Publicación

Años	Tipo de publicación			
	Revistas	Conferencias y talleres	Otros ²	Total
1998	1	1		2
1999	3	1	1	5
2000	1		1	2
2001	1		1	2
2002	1	2	1	4
2003	3		1	4
2004	3	3	1	7
2005	1	6	1	8
2006	2		1	3
2007	2	2		4
2008	3	1	1	5
2009	1	6	1	8
2010	1	1	1	3
2011	2	5	1	8
2012	2	1	1	4
Total	27	29	13	69

Resalta el hecho que desde el año 2004 hasta la fecha, es donde se concentra el mayor número de investigaciones, observándose un gran interés en los investigadores del área de la ingeniería de software en realizar estudios relacionados con las ARs y ALPs para LPS.

A continuación, un análisis más detallado se llevó a cabo en cada estudio incluido y el informe de la revisión.

C. Informe de la Revisión

En esta última fase, se presentan los resultados analíticos de la revisión sistemática. La extracción de los datos obtenidos y la síntesis del conocimiento considerando cada pregunta de investigación se discute a continuación.

1) *Con Relación a la Pregunta PII: ¿Cuáles métodos utilizan como estrategia el enfoque proactivo o el enfoque reactivo para el diseño de la AR y ALP?*

La Tabla VII resume los enfoques o técnicas que abordan el proceso de construcción de las arquitecturas desde la perspectiva del enfoque reactivo o *bottom-up*.

² Otros se refiere a los repositorios que tienen algunas universidades donde almacenan las investigaciones de sus estudiantes

Tabla VII: Construyen la Arquitectura (Enfoque Reactivo)

E ³	Método (autor(es)) (ref.)	TA ⁴
E1	Extra view to the QADA (Tan et al., 2012) [12]	ALP
E2	Framework RA aplicando SOA (Duro et al., 2005) [13]	AR
E3	RA based in Reverse Engineering (Grosskurth & Godfrey, 2005) [14]	AR
E4	RA basado en patrones (Mnñoz et al., 2005) [15]	AR
E5	ARES (Architecture Reference) approach [16]	AR
E6	Empirically-Grounded RA approach (Galster & Avgeriou, 2011) [17]	AR
E7	PuLSE-DSSA approach (Bayer et al., 2004) [18]	AR
E8	Prototype of reengineering approach (Jirapanthong, 2012) [19]	ALP
E9	Layered PLA approach (Bastarrica, 2006) [20]	ALP
E10	KobrA method (Anastasopoulos, 2002) [21]	ALP
E11	FAST (Family-Oriented Abstraction, Specification, and Translation) (Harsu, 2002) [22]	ALP
E12	An MDE-based PLA (Deng, 2009) [23]	ALP
E13	RA based in Prior Systems (Pinzger et al., 2004) [24]	AR
E14	Any-time variability approach (Hoek, 2004) [25]	ALP

La Tabla VIII resume los métodos que abordan el proceso de construcción de las arquitecturas desde la perspectiva del enfoque proactivo o *top-down*.

Tabla VIII: Construyen la Arquitectura (Enfoque Proactivo)

E ³	Método (autor(es)) (ref.)	TA ⁴
E15	Framework ACROSeT (Iborra et al., 2009) [26]	RA
E16	ERA (E-contracting Reference Architecture) approach (Angelov & Grefen, 2008) [27]	RA
E17	Interface-Activities Model (IAM) RA (Pollard & Duke, 2005) [28]	RA
E18	ProSA-RA (Nakagawa et al., 2009) [29]	RA
E19	A component-oriented approach centered on MDE (Panunzio, 2011) [30]	RA
E20	Cross-domain RA approach (Dobrica & Niemelä, 2008) [31]	RA
E21	ASRA (Agent System Reference Architecture) approach (Nguyen et al., 2010) [32]	RA
E22	Proactive Design Method for PLAs (Oizumi et al., 2012) [33]	ALP
E23	Decision-oriented approach (Dhungana, 2007) [34]	ALP
E24	SEPA Methodology (Barber & Graser, 1999) [35]	RA
E25	NEXOF-RA approach (Stricker et al., 2010) [36]	RA
E26	Process for Managing Variability (Kim et al., 2011) [37]	ALP
E27	UML profile approach for variations (Dobrica & Niemelä, 2007) [38]	ALP
E28	Style-specific approach (Lalanda, 1999) [39]	ALP
E29	Feature oriented approach (Zhu et al., 2007) [40]	ALP

En la Tabla IX, se resumen los resultados obtenidos en las Tablas VII y VIII, mostrando que no existe una tendencia o preferencia muy marcada en utilizar el enfoque proactivo o el reactivo, independientemente si es para diseñar la AR o ALP.

2) *Con Relación a la Pregunta PI2:* Identificar si los métodos utilizan algún patrón, estilo o modelo arquitectónico en particular como artefacto para la construcción de las ARs y ALPs, particularmente en lo referente a que artefactos reutiliza y adapta, solamente reutiliza o por el contrario agrega o incorpora algún nuevo artefacto en el proceso de construcción de estas arquitecturas.

En la Tabla IX se muestra la distribución de los estudios según el enfoque de construcción caracterizado por el tipo de arquitectura y asuntos de interés relacionados con las técnicas utilizadas por el método desde el punto de vista de la reutilización de artefactos arquitectónicos existentes.

Tabla IX: Distribución de Estudios por Enfoque de Desarrollo y Asuntos de Interés

Enfoque o modelo	Tipo de Arquitectura			Asuntos de interés en relación a artefactos utilizados en el método		
	AR ⁵	ALP ⁶	Total	Reutiliza y adapta	Reutiliza tal cual	Agrega artefacto
Proactivo	9	6	15	7	6	8
Reactivo	7	7	14	7	11	3
Total	16	13	29	14	17	11

Los artefactos son aquellas técnicas y/o elementos arquitecturales (estilo arquitectónico, patrón arquitectónico, patrón de diseño, etc.) que usan y/o implementan los métodos para la construcción de las arquitecturas.

Al realizar un análisis de las tendencias utilizadas en la construcción de las arquitecturas en los estudios analizados, según los datos suministrados por la Tabla IX, las siguientes preguntas fueron resueltas:

PI2.1. ¿Cuál es la forma de reutilización más utilizada en la construcción de estas arquitecturas?

Es evidente que la tendencia es la de reutilizar los artefactos existentes, en algunos casos se le realizan modificaciones a estos artefactos, en otros casos se incorporan en el método tal cual como originalmente fueron definidos.

PI2.2. ¿Cuál de los dos modelos (proactivo y reactivo) es el más utilizado en la construcción de estas arquitecturas?

La evidencia indica la tendencia en utilizar ambos enfoques, el proactivo y el reactivo por igual.

PI2.3. ¿Cuáles otros aspectos fueron evidenciados en los métodos analizados para la construcción de las arquitecturas?

La Tabla X muestra otras evidencias que reflejan el interés que tienen algunos autores, no solo en la construcción de las arquitecturas, sino también en clasificarlas, compararlas, evaluarlas y especificarlas formalmente.

Tabla X: Distribución de los Estudios por Criterios de Análisis

Criterios	TA		Total
	AR	ALP	
Construyen	16	13	29
Clasifican	3	1	4
Comparan	2	3	5
Evalúan	9	10	19
Especifican	0	3	3

Se destaca que los estudios [12][18][21][27][29][33][41], satisfacen dos de los criterios, particularmente describen un método para construir la arquitectura y luego la evalúan a través de características de calidad. Solo 1 estudio [42] mostró un enfoque para clasificar y evaluar las ARs.

³ Estudios

⁴ Tipo de arquitectura

⁵ Arquitectura de referencia

⁶ Arquitectura de líneas de productos

3) *Con Relación a la Pregunta "PI3": ¿Cuáles son las semejanzas y diferencias entre la AR y la ALP?*

La Tabla XI resume las similitudes y las diferencias entre ambas arquitecturas [4][12][14][16][17][19][30][31][36][43][44][45][46][47][48][49].

Tabla XI: Semejanzas y Diferencias entre AR y ALP

Semejanzas
Garantiza la interoperabilidad de los sistemas a través de la estandarización.
Facilita la creación de instancias de nuevas arquitecturas.
La flexibilidad y la interoperabilidad son dos de los requisitos no funcionales que guían su diseño.
Determinan el ámbito de la línea de productos que limita los productos que se pueden construir.
Se producen durante la actividad del diseño del dominio en la fase de ingeniería del dominio.
Comparten una arquitectura y un conjunto de componentes reutilizables entre una familia de productos.
Describe elementos básicos de la arquitectura de software para los sistemas que se derivan del mismo dominio.
Son arquitecturas de software que no contiene detalles de implementación.
Diferencias
La AR requiere la existencia de dominios bastantes maduros debido a su alto nivel de abstracción.
AR tiene mayor nivel de abstracción con respecto a la ALP.
Las ALP tienen que ver con la variabilidad entre los productos.
Las ALPs representan un grupo de sistemas que comparten elementos similares, pero producidos independientemente por una sola organización.
Las ARs representan el espectro visible y futuro (conjunto, gama, rango, variedad) de los sistemas en un dominio.
Las ALPs son menos abstractas que la RA, pero más abstractas que las arquitecturas concretas, es decir se sitúan en un nivel intermedio entre la AR y el producto concreto.
Las ALP son un tipo de arquitectura de referencia.
La ALP puede ser instanciada en una arquitectura de software mediante el refinado, extendiéndola, y en particular, especificando decisiones de implementación.
El rol de las ARs es capturar el conocimiento de las arquitecturas existentes. Sobre la base de la misión, visión y estrategias, y sobre las necesidades futuras de los clientes, y proporciona orientación a las múltiples organizaciones que evolucionan o crear nuevas arquitecturas.
La AR se crea mediante la captura de los elementos esenciales de las arquitecturas existentes y teniendo en cuenta las necesidades futuras y oportunidades, que van desde las tecnologías específicas, a los patrones de los modelos de negocio y segmentos de mercado.
La ALP especifica la descomposición de un sistema en subsistemas, la interacción entre subsistemas y la distribución de funcionalidad entre ellas.

IV. DISCUSIÓN

Los resultados de la revisión sistemática indican que, a pesar del gran interés en las AR y en las ALPs, la mayoría de los estudios aplican diferentes criterios en la forma o manera de concebir o diseñar la AR y las ALPs por lo cual ampliar la investigación en cuanto a la unificación de criterios en esta área sería de gran aporte en la comunidad de desarrolladores.

La evidencia indica que el término AR no se origina por el surgimiento de la metodología de ingeniería de líneas de productos de software como otro enfoque para la reutilización y la mantenibilidad de familias de productos de software, ya anteriormente se hablaba de las arquitecturas de referencias, esto se ve reflejado por los 25 estudios de los 69 que abordan las AR y no las relacionan con el SPL de forma explícita.

Las LPS es otro enfoque que hace énfasis en la reutilización de los activos de software presentes en la organización, la

preferencia es la de reutilizar los artefactos existentes (estilo arquitectónico, patrón, etc.), modificándolos o no, y en pocos casos incorporando nuevos artefactos en los métodos de desarrollo definidos en la organización.

De las políticas de desarrollo de software en la organización a nivel industrial, la toma de decisión de utilizar el enfoque proactivo o el reactivo dependerá de: si se quiere construir una nueva LPS sin considerar los productos existentes, modificar una LPS existente o reutilizar los productos existentes y de allí construir una nueva LPS.

Aunque hay autores que indican que la AR y la ALP son diferentes en cuanto al objetivo para el cual fueron concebidas, concretamente en el alcance del dominio determinado por la organización, en la mayoría de los estudios analizados, los autores utilizan ambos términos de forma equivalente.

V. CONCLUSIONES

La principal contribución de este trabajo fue presentar una perspectiva detallada sobre los enfoques de desarrollo proactivo y reactivo propuestos para la construcción de las AR y PLA, así como determinar la forma de reutilización de los artefactos en la construcción de estas arquitecturas.

La elección de parte de la organización del enfoque proactivo o reactivo como estrategia para el desarrollo de la arquitectura, dependerá de los tiempos de entrega de los productos y de la capacidad de inversión disponible en la organización; si la organización dispone de los recursos suficientes entonces tendrá la capacidad de abordar el enfoque proactivo que requiere gran cantidad de talento humano y de tiempo para diseñar la arquitectura; si la organización tiene tiempo en el mercado produciendo productos de software, el enfoque reactivo suele ser el elegido por esta, sin embargo se ha de tener en cuenta que los productos existentes no fueron inicialmente pensados o concebidos bajo el esquema de LPS, por lo cual el proceso de reingeniería necesario puede ser considerable, dependiendo de la madurez que exista en la organización en sus procesos de desarrollo de software y del uso de estándares de calidad en los productos presentes.

Existe una tendencia a la reutilización de artefactos existentes, como por ejemplo: estilos y patrones arquitectónicos, patrones de diseños, etc., que la incorporación de nuevos artefactos para la construcción de la AR y/o ALP.

Existe una carencia en el uso de modelos de calidad, por ejemplo el ISO 9126 o su actualización el ISO 25010, por lo cual una revisión documental sistemática que responda a la pregunta de cuáles estrategias y/o técnicas se han aplicado para proveer de calidad a las ARs y/o ALPs, y a partir de allí desarrollar un modelo de calidad para estas arquitecturas.

Esta revisión sistemática provee a la comunidad de desarrolladores de un insumo que apoye la toma de decisiones en la elección de utilizar el enfoque proactivo o reactivo para el abordaje de las LPSs como práctica organizacional para el desarrollo de sus productos de software, así como utilizar la AR o ALP como punto de partida para el diseño, dependiendo de la madurez que tenga la organización en cuanto a la metodología utilizada para el desarrollo de software y de la visión empresarial en relación a los futuros desarrollos.

REFERENCIAS

- [1] K. Pohl, G. Boeckle, and F. van der Linden, *Software Product Line Engineering - Foundations, Principles, and Techniques*, Springer Verlag, Berlin / Heidelberg, 2005.
- [2] E. Nakagawa, P. Oliveira, and M. Becker, *Exploring the Use of Reference Architectures in the Development of Product Line Artifacts*, in proceedings of the 15th International Software Product Line Conference, Vol 2, ACM, 2011.
- [3] B. Penzenstadler, *Evaluation and Testing of Approaches to Support Variants in Model-Based Design for Software Product Lines*, University of Passau, Thesis, 2006.
- [4] E. Nakagawa, P. Oliveira, and M. Becker, *Reference Architecture and Product Line Architecture: A Subtle but Critical Difference*, Software Architecture. Springer Berlin Heidelberg, pp. 207-211, 2011.
- [5] M. Guessi, L. B. R. de Oliveira, and E. Y. Nakagawa., *Representation of Reference Architectures: A Systematic Review*, Department of Computer Systems, University of São Paulo, SP, Brazil, 2011.
- [6] A. Rashid, J.C. Royer, and A. Rummler, *Aspect-Oriented Model-Driven Software Product Lines. The AMPLE Way*, Cambridge University Press, Cambridge, 2011.
- [7] C. W. Krueger, *Easing the Transition to Software Mass Customization*, in Software Product-Family Engineering, Springer Berlin Heidelberg, 2002, pp. 282-293.
- [8] B. Kitchenham, *Guidelines for Performing Systematic Literature Reviews in Software Engineering*, Version 2.3, EBSE Technical Report, Software Engineering Group, School of Computer Science and Mathematics, Keele University, UK and Department of Computer Science, University of Durham, UK, 2007.
- [9] B. Kitchenham, O. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, *Systematic Literature Reviews in Software Engineering, a Systematic Literature Review*, Information and software technology, vol. 51, no. 1, pp. 7-15, 2009.
- [10] B. Kitchenham, *Procedures for Performing Systematic Reviews*, Keele University and National ICT Australia Ltd, Tech. Report TR/SE-0401 and NICTA TR 0400011T.1, 2004.
- [11] B. Basem El-Haik and A. Shaout, *Software Design for Six Sigma: A Roadmap for Excellence*, John Wiley and Sons, 2010.
- [12] L. Tan, Y. Lin, and H. Ye. *Quality-Oriented Software Product Line Architecture Design*, Journal of Software Engineering and Applications, vol. 5, no. 7, pp. 472-476, 2012.
- [13] N. Duro, F. Moreira, J. Rogado, J. Reis, and Nestor Peccia, *Developing a Reference Architecture for the Ground Segment Software*, IEEEAC paper #1134, Version 1.14, IEEE, 2005.
- [14] A. Grosskurth and M. W. Godfrey, *A Reference Architecture for Web Browsers*, in proceedings of the 21st IEEE International Conference on Software Maintenance (ICSM'05), pp. 661-664, 2005.
- [15] J. Muñoz, J. Muñoz, F. J. Alvarez, and M. Romo, *Aplicación de un Método para Generar una Arquitectura de Referencia que Integre Sistemas Heredados y Bases de Datos*, Universidad Autónoma de Aguascalientes, México, 2005.
- [16] W. Eixelsberger, *Recovery of a Reference Architecture, A Case Study*, in proceedings of the third International Workshop on Software Architecture, ACM, pp. 33-36, 1998.
- [17] M. Galster and P. Avgeriou, *Empirically-Grounded Reference Architectures, A Proposal*, in proceedings of the joint ACM SIGSOFT conference--QoSA and ACM SIGSOFT symposium--ISARCS on QoSA and architecting critical systems--ISARCS, ACM, pp. 153-158, 2011.
- [18] J. Bayer, T. Forster, D. Ganesan, J. F. Girard, I. John, J. Knodel, R. Kolb, and D. Muthig, *Definition of Reference Architectures Based on Existing Systems, Lifecycle and Process for Family Integration*, Fraunhofer IESE-Report no. 2004, vol. 34.
- [19] W. Jirapanthong, *Experience on Re-Engineering Applying with Software Product Line*, International Journal of Computer Engineering Science (IJCES), vol. 2, no. 5, 2012.
- [20] M. C. Bastarrica, N. Hirschfeld-Kahler, and P. O. Rossel, *Product Line Architecture for a Family of Meshing Tools*, in Reuse of Off-the-Shelf Components, Springer Berlin Heidelberg, pp. 403-406, 2006.
- [21] M. Anastasopoulos, C. Atkinson, and D. Muthig, *A Concrete Method for Developing and Applying Product Line Architectures*, in Objects, Components, Architectures, Services, and Applications for a Networked World, Springer Berlin Heidelberg, pp. 294-312, 2003.
- [22] M. Harsu, *FAST Product-Line Architecture Process*, Tampere University of Technology, 2002.
- [23] G. Deng, D. C. Schmidt, A. Gokhale, J. Gray, Y. Lin, and G. Lenz, *Evolution in Model-Driven Software Product-Line Architecture*, Designing Software-Intensive Systems: Methods and Principles, 2008.
- [24] M. Pinzger, H. Gall, J. F. Girard, J. Knodel, C. Riva, W. Pasman, C. Broerse, and J. G. Wijnstra, *Architecture Recovery for Product Families*, in Software Product-Family Engineering, Springer Berlin Heidelberg, pp. 332-351, 2004.
- [25] A. van der Hoek, *Design-Time Product Line Architectures for Any-Time Variability*, Science of Computer Programming, vol. 53, no. 3, pp. 285-304, 2004.
- [26] A. Iborra, D. Alonso, F. J. Ortiz, J. Pastor, P. Sánchez, and B. Álvarez, *Design of Service Robots - Experiences using Software Engineering*, Robotics & Automation Magazine, IEEE, vol. 16, no. 1, pp. 24-33, 2009.
- [27] S. Angelov and P. Grefen, *An E-Contracting Reference Architecture*, Journal of Systems and Software, vol 81, no. 11, pp. 1816-1844, 2008.
- [28] J. Pollard and R. Duke, *A Reference Architecture for Instructional Educational Software*, in ACM International Conference Proceeding Series, pp. 43-52, 2005.
- [29] E. Y. Nakagawa, E. F. Barbosa, and J. C. Maldonado, *Exploring Ontologies to Support the Establishment of Reference Architectures, An Example on Software Testing*, in European Conference on Software Architecture WICSA/ECSA 2009, pp. 249-252, 2009.
- [30] M. Panunzio, *Definition, Realization and Evaluation of a Software Reference Architecture*, Technical Report UBLCS-2011-07, Department of Computer Science, University of Bologna, Italy, 2011.
- [31] L. Dobrica and E. Niemelä, *An Approach to Reference Architecture Design for Different Domains of Embedded Systems*, in Software Engineering Research and Practice, pp. 287-293, 2008.
- [32] D. N. Nguyen, K. Usbeck, W. M. Mongan, C. T. Cannon, R. N. Lass, J. Salvage, W. C. Regli, I. Mayk, and T. Urness, *A Methodology for Developing an Agent Systems Reference Architecture*, in Agent-Oriented Software Engineering XI, Springer Berlin, pp. 177-188, 2011.
- [33] W. N. Oizumi, A. C. Contieri, G. G. Correia, T. E. Colanzi, S. Ferrari, I. M. S. Gimenes, E. A. Oliveira, A. F. Garcia, and P. C. Masiero, *On the Proactive Design of Product-Line Architectures with Aspects, An Exploratory Study*, in proceedings of the 36th Computer Software and Applications Conference (COMPSAC 2012), pp. 273-278, 2012.
- [34] D. Dhungana, R. Rabiser, and P. Grünbacher, *Decision-Oriented Modeling of Product Line Architectures*, in the 6th Working IEEE/IFIP Conference of Software Architecture (WICSA'07), Mumbai, India, January 2007.
- [35] K. S. Barber, T. Graser, J. Holt, and C. Silva, *Representing Domain Reference Architectures by Extending the UML Metamodel*, in 20th International Conference on Software Engineering and Knowledge Engineering, San Francisco, USA, July 1999.
- [36] V. Stricker, K. Lauenroth, P. Corte, F. Gittler, S. De Panfilis, and K. Pohl, *Creating a Reference Architecture for Service-Based Systems: A Pattern-Based Approach*, in proceeding of Towards the Future Internet - Emerging Trends from European Research, 2010.
- [37] Y. G. Kim, S. Kee Lee, and S. B. Jang, *Variability Management for Software Product-Line Architecture Development*, International Journal of Software Engineering and Knowledge Engineering, vol. 21, no. 07, pp. 931-956, 2011.
- [38] L. Dobrica and E. Niemelä, *Modeling Variability in the Software Product Line Architecture of Distributed Middleware Services*, VTT Technical Research Centre of Finland, Oulu, Finland, 2007.
- [39] P. Lalanda, *Style-Specific Techniques to Design Product-Line Architectures*, Thomson-CSF Corporate Research Laboratory, Domaine de Corbeville, Orsay, France, 1999.
- [40] C. Zhu, Y. Lee, W. Zhao, and J. Zhang, *A Feature Oriented Approach to Mapping from Domain Requirements to Product Line Architecture*, in Software Engineering Research and Practice, pp. 219-225, 2007.
- [41] J. Bayer, O. Flege, and C. Gacek, *Creating Product Line Architectures*, in Software Architectures for Product Families, Springer Berlin Heidelberg, pp. 210-216, 1999.
- [42] C. Maga and N. Jazdi, *A Survey on Determining Factors for Modeling Reference Architectures*, in proceedings of OOPSLA, Florida, USA, 2009.
- [43] G. Muller and P. van de Laar, *Researching Reference Architectures and their Relationship with Frameworks, Methods, Techniques, and Tools*, 7th Annual Conference on Systems Engineering Research (CSER 2009), S08-46, Loughborough University, UK, 2009.

- [44] P. Avgeriou, *Describing, Instantiating and Evaluating a Reference Architecture: A Case Study*, Enterprise Architect Journal, 2003.
- [45] B. Graaf, H. van Dijk, and A. van Deursen, *Evaluating an Embedded Software Reference Architecture - Industrial Experience Report*, in proceedings of the 9th European Conference of Software Maintenance and Reengineering (CSMR 2005), pp. 354-363, 2005.
- [46] U. Eklund, Ö. Askerdal, J. Granholm, A. Alminger, and J. Axelsson, *Experience of Introducing Reference Architectures in the Development of Automotive Electronic Systems*, 2005.
- [47] S. Giesecke, W. Hasselbring, and C. von Ossietzky, *Taxonomy of Architectural Style Usage*, in proceedings of the 2006 conference on Pattern Languages of Programs, ACM, 2006.
- [48] B. P. Gallagher, *Using the Architecture Tradeoff Analysis Method to Evaluate a Reference Architecture: A Case Study*, CMU/AEI-2000-TN-007, Carnegie-Mellon University, Pittsburgh Software Engineering Institute, 2000.
- [49] J. Bayer, D. Ganesan, J. F. Girard, J. Knodel, R. Kolb, and K. Schmid, *Definition of Reference Architectures based on Existing Systems*, Fraunhofer IESE-Report, vol. 34, 2004.

Desarrollo de Aplicaciones Usando Modelos de Procesos Flexibles

Maria Guevara, Dinarle Ortega, Argenis Guerrero
maigualidaguevara@gmail.com, dinarleortega@gmail.com, ajguerrero.88@gmail.com

Departamento de Computación, Universidad de Carabobo, Valencia, Venezuela

Resumen: La presente investigación aborda el desarrollo de una Aplicación de Flujo de Trabajo usando un Modelo de Procesos Flexibles, el cual promueve mayor facilidad y menor tiempo para plasmar durante el diseño del proceso, los cambios en respuesta a las variaciones del entorno, utilizando patrones workflow y algunos métodos para el modelado de la flexibilidad, empleando el estándar BPMN (Business Process Modeling Notation). Se utiliza como estudio de caso el proceso “Gestión del Requisito Obligatorio Apreciación Cultural de los planes de estudio de pregrado de la FaCyT (Facultad de Ciencia y Tecnología)”. La metodología de desarrollo de la Aplicación es una instanciación de RUP incorporando algunos aspectos de BMM (Business Modeling Method).

Palabras Clave: Modelos de Procesos Flexibles; Flexibilidad en Tiempo de Diseño; BPMN; Aplicación de Flujo de Trabajo.

Abstract: This research addresses the development of a Workflow application using a Flexible Process Model, which promotes an easier and less time consuming way to translate required changes into the model, as response to the environment variations during the design process, using workflow patterns and some methods for flexibility modeling with the BPMN (Business Process Modeling Notation) standard. It is used as a case study the process “Management of the Cultural Appreciation Mandatory Requirement of the FaCyT undergraduate curricula.” The development methodology is a RUP instantiation, incorporating some aspects of BMM (Business Modeling Method).

Keywords: Flexible Process Models; Design Time Flexibility; BPMN; Workflow Application.

I. INTRODUCCIÓN

Un problema común en los proyectos de desarrollo de software es la inconformidad de los usuarios respecto al producto final; el mismo no es capaz de reflejar la realidad organizativa ni brindar soporte a las actividades como se espera. Diversas causas influyen: los requisitos no han sido bien capturados, no se ha entendido el problema, o la complejidad y dinámica de las reglas de negocio obliga a una constante readaptación del software.

Por otro lado, en la actualidad se ha extendido el enfoque de desarrollo de software partiendo de procesos de negocio bien definidos [1][2]. De ahí, la importancia de establecer con detalle los procesos de negocio en los cuales la organización sustenta su adecuado funcionamiento. Es por ello, que la disciplina de modelado de procesos de negocio ha captado atención en la comunidad de investigación de desarrollo de software, ya que asume como parte de sus premisas, que al disponer de una especificación o modelo de proceso, es posible obtener un conjunto importante de requisitos para el desarrollo de aplicaciones de software, de igual forma es la oportunidad de realizar transformaciones sucesivas del modelo, hasta

alcanzar un modelo de proceso ejecutable, con lo cual se apresura también el desarrollo [1][3][4].

Los procesos de negocio dentro de la organización evolucionan a través de los años. Adicionalmente, los entornos de las organizaciones se han vuelto más complejos y cambiantes [5], siendo deseable que la especificación de sus procesos contemple características ó elementos propios de flexibilidad para tener capacidad de adaptación efectiva ante necesidades imprevistas.

De modo que, en este trabajo se investigó sobre cómo especificar modelos procesos de negocio flexibles expresados con el estándar BPMN (Business Process Modeling Notation), que puedan ser interpretados por herramientas para la gestión de procesos de negocio o BPMS (Business Process Management Suite), y de esta manera agilizar el desarrollo de sistemas basados en modelos de procesos. Se utilizó como estudio de caso el proceso “Gestión del Requisito Obligatorio (RO) Apreciación Cultural de los planes de Estudio de FaCyT” de la Universidad de Carabobo, dada su importancia para la administración de los Planes de Estudios y la naturaleza variable de aspectos que lo afectan.

El modelo de proceso flexible expresado en BPMN obtenido en esta investigación, constituye la base para el desarrollo de una Aplicación de Flujo de Trabajo que le brinda soporte automatizado, incluyendo abstracciones para responder a los cambios frecuentes. Para esto, se utilizan algunos patrones workflow para su realización [6] y métodos para el modelado de la flexibilidad [7].

Para guiar el desarrollo de la aplicación de flujo de trabajo a partir del modelo de proceso flexible de Gestión del RO Apreciación Cultural, se instancia RUP (Rational Unified Process) [8] incorporando algunos aspectos de BMM (Business Modeling Method) [9]. Se emplea El BPMS Bonita Open Solution [10] del cual destacamos el ambiente para obtener un modelo de proceso ejecutable y el motor de flujo de trabajo para su ejecución.

II. MODELOS DE PROCESOS DE NEGOCIO FLEXIBLES

Un proceso de negocio es una composición de actividades que, de forma conjunta, colaboran para alcanzar un objetivo común [11]. Adicionalmente, si se desea que éste responda de manera ágil en un entorno cambiante, se debe incluir algunos elementos de flexibilidad en su definición [2].

El término de flexibilidad en los procesos de negocio es mencionado frecuentemente, siendo difícil de comunicar en términos concretos y cuantificables lo que realmente es y qué implica [4][12]. Son muchas las definiciones, y significativas las investigaciones sobre técnicas para su manejo, por ende, dependiendo de la definición asumida, así será manejada y modelada la flexibilidad [13].

La flexibilidad en un proceso, podría expresarse como la adaptación a los cambios en función de tres dominios [12]:

- Tiempo: realizarlo más rápidamente.
- Costo: disminuir la utilización de los recursos.
- Facilidad: responder a los cambios con menor esfuerzo.

El dominio de la Facilidad se puede formular en términos de: cambiar el menor número de elementos, lograr una modularidad adecuada donde el impacto de los cambios se focalice e incorporar nuevos requisitos con poco esfuerzo y rapidez.

El modelo de proceso de negocio flexible que se elabora en esta investigación, está basado en los dominios de Facilidad y Tiempo. En la Figura 1, de elaboración propia, se muestra los tres dominios en función de los cuales se puede expresar la flexibilidad; el área sombreada indica los dominios contemplados.

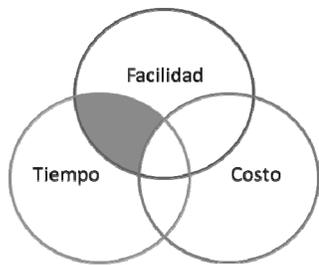


Figura 1: Dominios para Expresar la Flexibilidad en los Procesos

De igual importancia, se supone establecer la definición del término procesos de negocio flexible, la cual es producto de la consideración de diversas investigaciones [14][15]: Un proceso de negocio es flexible, si tiene la capacidad de cambiar sin ser reemplazado completamente, es decir, alterar una sección o parte del proceso con el menor impacto, esfuerzo y tiempo, cuando en la organización se producen cambios operacionales y/o funcionales.

Los modelos de Procesos de Negocio y en particular, los flexibles se pueden expresar usando notaciones estándares como BPMN porque: (1) provee los elementos para expresar situaciones o variaciones que afectan el flujo normal del proceso, a través de la especificación de caminos alternativos de ejecución, y (2) dispone de la capacidad para modelar eventos y puntos de decisión exclusivos para la diversificación del flujo.

Adicionalmente, se debe utilizar técnicas que facilitan las abstracciones de la flexibilidad, es decir, elementos sintácticos y constructos de BPMN, patrones, métodos y principios de diseño.

A. Flexibilidad en Tiempo de Diseño

La flexibilidad ha sido estudiada tanto en términos de los factores que la motivan como de las formas para lograrla en los procesos de negocio. Se ha clasificado en su mayoría desde una perspectiva del control del flujo [6], proponiendo distintas opciones de realización. Puede definirse en tiempo de Diseño ó en tiempo de Ejecución, dependiendo de la naturaleza del entorno en el que se desenvuelve el proceso, por lo cual demanda comprender las variaciones y perturbaciones que podrían afectarlo.

En el caso de la flexibilidad en tiempo de diseño las variaciones pueden anticiparse y pre-definirse; por tanto, la respuesta a la flexibilidad se plasma en el modelo del proceso expresado con BPMN. Estudiando las variaciones en el entorno se ha de proporcionar alternativas de ejecución utilizando variantes del constructo de elección para la especificación de caminos alternativos. Así que, se requiere identificar las variaciones de forma exhaustiva, definir las claramente, que sean mutuamente excluyentes y que los puntos de decisión estén perfectamente representados [14].

No siempre, se puede prever todas las posibilidades de ejecución, y tratar de describirlas totalmente, incrementa la complejidad del modelo resultante [6]. En este sentido, el uso de algunos patrones workflow constituye opciones de realización para modelar la flexibilidad en tiempo de diseño [6], estos son:

- Paralelismo: ejecutar un conjunto de tareas en paralelo.
- Elección: seleccionar una o más tareas para la ejecución posterior.
- Iteración: ejecutar repetidamente una tarea.
- Entrelazado: ejecutar cada tarea, de un conjunto de ellas, en cualquier orden.
- Múltiples instancias: múltiples instancias concurrentes de una tarea.
- Cancelación: retirar una tarea en ejecución.

Además, para el modelado de este tipo de flexibilidad, se precisaron un conjunto de métodos con base en los principios de granularidad y modularidad. Estos métodos consideran que, la aplicación razonable de los principios mencionados, reduce considerablemente el Tiempo y aumenta la Facilidad para reflejar los cambios. Con la granularidad (nivel de detalle) se asegura que el análisis de impacto de los cambios y su verificación, sea más rápido y sencillo, ya que es posible visualizar en detalle una parte específica del proceso; y con la modularidad, se asegura que los cambios necesarios no afecten al proceso completo.

B. Métodos para el Modelado de la Flexibilidad

Esta sección incluye los métodos investigados y aplicados para el modelado de flexibilidad.

1) *Jerarquía de Procesos*: Al establecer cierta jerarquía para el proceso se promueve los beneficios de utilizar los principios de granularidad y modularidad. Así, el modelo de proceso estará organizado jerárquicamente en tres niveles: proceso de negocio, flujo de trabajo del proceso (secuencia detallada de las actividades del proceso especificado en el primer nivel) y procedimientos/tareas del negocio (procedimientos necesarios para ejecutar las actividades en el flujo de trabajo del proceso). La Figura 2 ilustra los niveles [7].

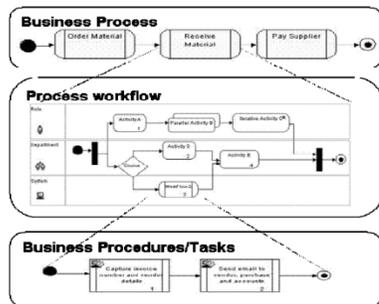


Figura 2: Jerarquía de Procesos

2) *Elementos Condicionales*: disminuir el uso de las rígidas rutas predefinidas, y por el contrario, proporcionar varias rutas y alternativas de ejecución haciendo uso de elementos condicionales expresados en términos de constructos básicos para el control, divergencia y convergencia de la secuencia. De esta manera se incorpora flexibilidad en el modelo de proceso.

3) *Reglas del Negocio*: las Reglas del Negocio son una colección de políticas aplicadas en una organización y que rigen su funcionamiento, restringiendo, condicionando y/o controlando el dominio en el que opera el proceso y su ejecución. La variabilidad en dichas Reglas incide en la necesidad de diseñar un proceso flexible. Es por ello, que son implementadas como decisiones IF/THEN, definidas usando conectores (secuencia, división paralela, elección exclusiva). La incorporación de decisiones infunde el comportamiento al modelo del proceso. Igualmente, la consideración de las Reglas del Negocio permite la modelación del segundo nivel de la jerarquía de procesos.

4) *Normas*: reflejan las prácticas y/o tareas dentro de una organización, así como, las regularidades en el

comportamiento de los miembros de la misma que coordinan sus acciones.

La aplicación de los métodos enumerados, permite la captura de cuatro dimensiones diferentes de flexibilidad, señaladas como las formas para abordar el problema de flexibilidad desde una perspectiva de modelado en tiempo de diseño.

- Flexibilidad en la secuencia del proceso: Facilidad de cambio en la secuencia o programación de las actividades.
- Flexibilidad de las reglas de negocio aplicables: cambios en la normativa, las reglas y regulaciones de negocios, la adaptación de las nuevas políticas.
- Prácticas flexibles: Capacidad de incorporar nuevas prácticas y métodos de trabajo.
- Flexibilidad en el manejo de excepciones de procesos: Facilidad de elegir soluciones que más se corresponden con el problema y el contexto actual.

Las dimensiones asumidas en esta investigación y los métodos aplicados se indican en las celdas grises de la Tabla I tomada de [7].

Tabla I: Dimensiones de Flexibilidad Consideradas en la Investigación y los Métodos de Modelado Aplicados

Dimensiones de flexibilidad	Métodos de modelado
Flexibilidad en la secuencia del proceso	Jerarquía de procesos. Elementos condicionales.
Flexibilidad en las reglas del negocio	Captura de las reglas del negocio
Prácticas flexibles	Tareas del negocio. Normas.
Flexibilidad en el manejo de excepciones (anticipadas)	Elementos condicionales, Reglas Normas

C. BPMN y la Flexibilidad

BPMN [16] se utiliza como notación para la especificación dada su capacidad para modelar flexibilidad en los procesos de negocio. En este sentido resaltan los siguientes aspectos: (1) permite expresar distintos niveles de abstracción (hasta el nivel inferior de tareas), esencial para establecer la jerarquía de procesos; (2) los constructos *Eventos* y *Gateways*, su amplia diversidad, la semántica que indica su comportamiento, entre otros, permiten modelar las opciones de realización mencionadas (paralelismo, entrelazado, cancelación, etc.) propuestas por [6], en tiempo de diseño y abarcar adecuadamente las dimensiones de flexibilidad seleccionadas; (3) la existencia de una amplia gama de herramientas/entornos BPMs (Business Process Management) que permiten la transformación del modelo BPMN para una subsecuente implementación, tal es el caso de Bonita Open Solution.

En cuanto a los Eventos y Gateways se precisa lo siguiente:

- Eventos intermedios: para indicar cuando algo sucede/ocurre después de que un Proceso ha comenzado y antes de que haya finalizado.
- Gateways basado en eventos: para representar un punto de bifurcación alternativo donde la decisión está basada en

Eventos que pueden ocurrir, en vez de condiciones orientadas a datos.

III. METODOLOGÍA DE DESARROLLO DE SOFTWARE USANDO RUP Y BMM

En esta investigación se utilizó una instanciación de RUP complementada con actividades y artefactos del método BMM, para guiar el desarrollo de la aplicación de flujo de trabajo con un enfoque basado en procesos. Específicamente se adecuó la disciplina Modelado del Negocio de RUP con el método BMM, de donde resalta el artefacto BPD (Business Process Diagram) elaborado o Modelo de Proceso Flexible, expresado en BPMN, el cual constituye una base para la extracción de los requisitos del sistema. Asimismo, dicho modelo es posteriormente refinado y transformado en un Modelo de Proceso Ejecutable empleando funcionalidades del BPMS (BPM Suite) Bonita Open Solution.

Dado que, el alcance de este trabajo es desarrollar una aplicación para brindar soporte al proceso Gestión del RO (Requisito Obligatorio) Apreciación Cultural bajo un enfoque flexible, más no es, realizar un modelo de negocio completo de la organización Coordinación de Cultura (describir todos los procesos que coexisten en ella), el método BMM es instanciado de manera que incluya los aspectos de interés para el problema.

En la Tabla II se presenta las actividades y artefactos más significativos de la instanciación de RUP complementado con BMM.

IV. DESARROLLO DE LA APLICACIÓN

Esta sección describe la experiencia de desarrollo de la aplicación; en particular se proporcionan detalles de algunas actividades y artefactos de las fases de Inicio, Elaboración y Construcción, que abarcan aspectos significativos para el diseño del modelo de proceso flexible y su transformación en un Modelo de Proceso Ejecutable (o Aplicación de Flujo de Trabajo), por razones de espacio.

En este punto, es oportuno indicar las razones por las cuales se estudia el proceso “Gestión del Requisito Obligatorio Apreciación Cultural de los planes de estudio de pregrado de la FaCyT”. La Facultad está conformada por un conjunto de procesos administrativos, académicos, entre otros. En cuanto a los académicos el proceso en estudio está sujeto a cambios frecuentes, porque la Coordinación de Cultura encargada de su administración, continúa en búsqueda de captar el interés y lograr una participación más activa del estudiante, por tanto realiza ajustes frecuentes en su planificación, desarrollo, y modificaciones en los mecanismos de evaluación. Así mismo, las habituales reformas en los planes de estudio de pregrado de la FaCyT y el variable número de estudiantes con opción de cursar el RO, son razones que aunadas a las antes nombradas, demandan cambios.

A. Fase de Inicio

Esta subsección presenta algunos de los artefactos elaborados.

Expresión del Dominio de la Organización y Contexto del Proceso.

Apreciación Cultural engloba un conjunto de actividades para la promoción cultural dentro de la facultad, dichas actividades favorecen la realización satisfactoria del RO a la vez de incentivar el interés por la cultural e incorporarla al quehacer académico de manera integral.

Nombre, ubicación y descripción de la organización.

La Coordinación de Cultura es una unidad adscrita a la Dirección de Extensión de la FaCyT que tiene como misión la planificación, desarrollo y difusión del quehacer cultural en la misma para el logro de una participación activa por parte del estudiantado en el proceso de formación académica.

Funciones de la Organización.

La Coordinación de Cultura tiene entre sus funciones:

- Preparación de la logística para eventos especiales de carácter artístico y cultural.
- Planificación y desarrollo de actividades de formación y de manifestación cultural a través de talleres, jornadas, eventos culturales, exposiciones y demás actividades de índole cultural.
- Coordinación y organización de eventos culturales de la Facultad.
- Desarrollo de programas encaminados a la promoción y divulgación de las actividades culturales de Facultad.
- Apoyo en la realización de eventos culturales, artísticos y/o deportivos programados por los Departamentos y la Unidad de Formación Socio-Humanística.

Modelado de Fines y Objetivos de la Organización

El modelo de objetivos se diseñó a partir del análisis de las funciones actuales de la Coordinación de Cultura, y de las unidades o direcciones superiores a las cuales ésta obedece, por cuanto no existe documentación institucional. El conjunto de objetivos formulados incluye la gestión del RO de Apreciación Cultural. La Figura 3 muestra parte del árbol de objetivos.

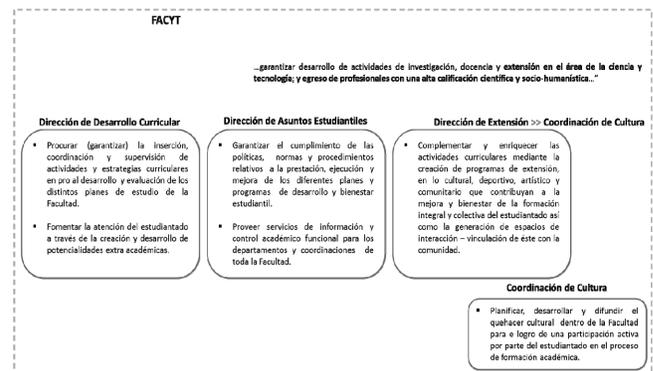


Figura 3: Árbol de Objetivos de la Coordinación de Cultura de la FaCyT

Tabla II: Instanciación de RUP y BMM

INICIO		ELABORACIÓN		CONSTRUCCIÓN		TRANSICIÓN	
MODELADO DEL NEGOCIO		MODELADO DEL NEGOCIO		IMPLEMENTACIÓN		PRUEBAS	
Actividad	Artefacto	Actividad	Artefacto	Actividad	Artefacto	Actividad	Artefacto
Expresión del dominio de organización y contexto del proceso.	Documento textual sobre la realidad de organización y el contexto del proceso.	Registro de nuevos escenarios identificados que denotan flexibilidad.	Lista actualizada de escenarios Identificados	Programación <i>Java</i> actividades proceso y Generación archivos .bar (Un Subproceso/iteración)	Modelo Proceso ejecutable	Elaboración de manual de usuario.	Manual de Usuario para navegación dentro de Aplicación.
				Programación <i>Bonita Solution</i> Interfaz Usuario: formularios y generación archivos .war (Un Subproceso/iteración)	Interfaz Gráfica de Usuario		
		Determinación constructos BPMN que expresen flexibilidad	Constructos de modelado BPMN que expresen flexibilidad	Conexión a Base Datos	Enlaza Base Datos My SQL		
				Creación entorno de ejecución del proceso	Paquete Tomcat: servidor web, motor ejecución Bonita y archivos .war		
Modelado de fines y objetivos de la organización.	Árbol de objetivos.	Modelado de proceso Flexible	BPD Flexible			Realización pruebas de flexibilidad.	Evaluación de flexibilidad usando opciones realización [6]
		REQUISITOS					
		Refinamiento del Documento de Requisitos.	Documento textual de requisitos refinado.			Aplicación métricas para evaluación mantenibilidad [17].	Tabla comparativa versiones modelo para apreciar variabilidad de característica mantenibilidad
Especificación de reglas.	Descripción y categorización de reglas del negocio que rigen a la organización y al proceso	ANÁLISIS Y DISEÑO					
		Actividad	Artefacto				
		Definición de Arquitectura candidata.	Documento Arquitectura Vista Lógica: BPD Flexible refinado nivel procedimientos y tareas (Un subproceso por iteración)				
			Documento Arquitectura Vista Usuarios				
			Documento Arquitectura Vista Implementación				
Vista Datos: Diagrama Entidad-Relación y Modelo Relacional							
Modelado de los procesos.	Lista de Subprocesos identificados Modelo BPMN resaltando subprocesos con técnica coloración.						
Identificación de objetos.	Lista elementos concretos o abstractos circulan entre procesos						
Identificación de Actores.	Descripción de Actores						
Especificación BPMN del proceso. Aplica principios: granularidad y modularidad, método jerarquía de procesos	Diagrama del proceso (BPD).						
REQUISITOS							
Actividad	Artefacto						
Elaboración Documento Requisitos derivados BPD	Documento textual de Requisitos.						

Especificación de Reglas

Las reglas de negocio provienen de la revisión de las bases para el cumplimiento de Apreciación Cultural y de la Normativa de Evaluación de los aprendizajes para los Estudiantes de la FaCyT.

Bhat y Deshmukh [7], proponen una categorización de Reglas del Negocio: Ejecución de proceso, Eventos de Negocios, Reglamentos-y-políticas y Limitaciones; por tanto, se procuró precisar las reglas que rigen el proceso en estudio bajo esta clasificación, a fin de abordar apropiadamente la dimensión de Flexibilidad de las Reglas del Negocio. Además, al aplicar la mencionada categorización, mejoró la comprensión de los criterios de decisión ó condiciones que definen y controlan al proceso durante su ejecución, y proporcionar claridad al momento de plasmar flexibilidad en la secuencia del proceso. Asimismo, asintió el conocimiento sobre la posible modelación del segundo nivel de jerarquía de procesos (flujo de trabajo del proceso).

En vista de que la mayoría de las reglas, constituyen el conjunto de bases establecidas para que el estudiante pueda cumplir satisfactoriamente con el RO, éstas han sido ubicadas principalmente en la categoría de Ejecución de proceso, y modeladas como decisiones para establecer así diversos enrutamientos.

Identificación de Actores

Se analizó cómo la Coordinación de Cultura ejecuta las actividades para dar cumplimiento a lo contemplado en la normativa vigente. De ahí se determinó, que el RO Apreciación Cultural está dirigido al estudiante, quien es el actor principal. La Tabla III contiene los Actores y su breve descripción.

Identificación de Subprocesos

Partiendo de una versión del Modelo de proceso expresado en BPMN, se procede a resaltar en colores la agrupación y relación de actividades que conforman los subprocesos, entre los distintos pools (actores), a lo largo del proceso; con el fin de, facilitar su visualización y apreciar el nivel de abstracción correspondiente a subprocesos, Tabla IV, de conformidad con el método de jerarquía de procesos.

Tabla III: Identificación de Actores del Proceso de Gestión del RO Apreciación Cultural

ACTOR	ROL / DESEMPEÑO: Competencias, habilidades y actitudes:	
Estudiante regular	Inscribirse en cualquiera de los distintos talleres ofertados por la Coordinación de Cultura para dar cumplimiento al RO, recibir notificaciones sobre avances en el cumplimiento del RO, llevar un seguimiento remoto de las actividades.	
Coordinación	Coordinador de Cultura	Vigilar el desarrollo de los talleres ofertados, mediante la generación y análisis de distintos reportes, para visualizar en cualquier momento el estado en que se encuentran los talleres y el estudiantado asistente. Llevar el control de los estudiantes que tienen la obligación de cursar el RO como parte de su plan de estudio así como de los que se han registrado en los distintos talleres.
	Asistente Cultural	Administración, transcripción y verificación de las listas de asistencias recibidas semanalmente (generadas en cada clase dada por parte de un taller), contribuye a minimizar el tiempo dedicado para llevar un control de gestión actualizado del RO.
	Beuario	Asistir a la Coordinación de Cultura en todo lo referente a las actividades que esta realiza con el fin de promover el interés cultural dentro de la facultad. Logística.
Instructor	Dictar los talleres y llevar a cabo la captura de las asistencias de los estudiantes a las clases. Proponer y ejecutar el mecanismo de evaluación final a los estudiantes y dictamina si éste ha aprobado o no.	
Director de Extensión	Ejecutar la política de extensión en la Facultad diseñando planes, programas y proyectos en materia de extensión, para el desarrollo, difusión y promoción de actividades de índole cultural para la comunidad de la FaCyT. Asigna espacios para el desarrollo y manifestación de los talleres dentro de la misma. Supervisar el desarrollo del requisito Apreciación Cultural.	

Tabla IV: Subprocesos Identificados

ORDEN	SUBPROCESO GLOBAL (A NIVEL DEL PROCESO)	SUBPROCESO (A NIVEL DEL POOL)
1	Planificación del Requisito Obligatorio A.C	<ul style="list-style-type: none"> ▪ Planificar Módulo (Coordinación)
2	Apertura y ajuste del Requisito Obligatorio A.C	<ul style="list-style-type: none"> ▪ Registrar taller (estudiante) ▪ Ajustar planificación (Coordinación)
3	Desarrollo y seguimiento del Requisito Obligatorio A.C	<ul style="list-style-type: none"> ▪ Asistir a clases (estudiante) ▪ Supervisar el desarrollo del Requisito Obligatorio Apreciación Cultural (Coordinación) ▪ Dictar clases (Instructor)
4	Finalización y cierre del Requisito Obligatorio A.C	<ul style="list-style-type: none"> ▪ Organizar presentación final (Coordinación) ▪ Participar en presentación final (Estudiante) ▪ Dirigir presentación final (Instructor)

B. Fase de Elaboración

Constructos de Modelado BPMN que Expresen Flexibilidad

El enfoque de flexibilidad establecido, en tiempo de diseño, se basa en una combinación de tres aspectos: (1) reglas del negocio concebidas como decisiones, descritas en una sintaxis muy cercana al lenguaje natural y separadas lo más posible de los elementos gráficos y de datos persistentes para facilitar los cambios, (2) eventos intermedios (las diversas situaciones que

podrían eventualmente ocurrir, se representan en términos de este constructo) y (3) gateway's (compuertas lógicas para el control de flujo); creando así dentro del modelo BPMN, secciones de flujos de secuencia consideradas flexibles que describen cómo se responde a una variación percibida con anticipación.

En particular, el uso de los distintos tipos de eventos intermedios y gateway's, permite representar flexibilidad en cuanto al control del flujo (divergencia, convergencia, caminos de excepción), además son idóneos para la creación de una serie de caminos alternativos.

Definición de Arquitectura Candidata: Documento de Arquitectura Vista Lógica

La Figura 4 muestra un segmento del Modelo de Proceso Flexible Gestión del RO, se observan tres pools: Estudiante, Coordinación de Cultura e Instructor; también, se distinguen porciones coloreadas y numeradas del BPD, que indican opciones de realización de van der Aalst [6], los métodos de modelado de flexibilidad utilizados y los constructos de la notación BPMN para flexibilizar el modelo de proceso. Por ejemplo:

El Modelo de Proceso Flexible Gestión del RO, está conformado por tres pools: Estudiante, Instructor y Coordinación de Cultura; las Figuras 4 y 5, resaltan porciones numeradas del tercer pool nombrado, indicando opciones de realización, los métodos de modelado de flexibilidad utilizados y los constructos de la notación BPMN empleados. Por ejemplo:

1. Variabilidad en las reglas que definen y rigen la ejecución del proceso, evitando usar valores estáticos y garantizando que puedan ser modificados en cualquier punto del flujo.
2. Uso de compuertas AND para garantizar paralelismo; esto es, ejecución simultanea de tareas o secciones específicas del proceso.
3. Establecimiento de tareas bucles; es decir, realización de una tarea tantas veces se indique.
4. Bifurcación exclusiva del flujo para ejecutar una o varias tareas específicas entre otras más.

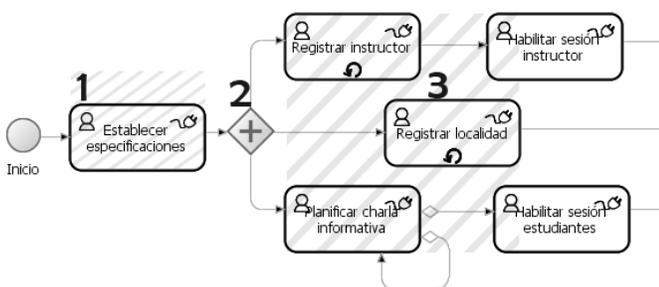


Figura 4: Segmento del Modelo Flexible Señalando con 1, 2 y 3, Variabilidad en las Reglas, Constructo AND y Bucles

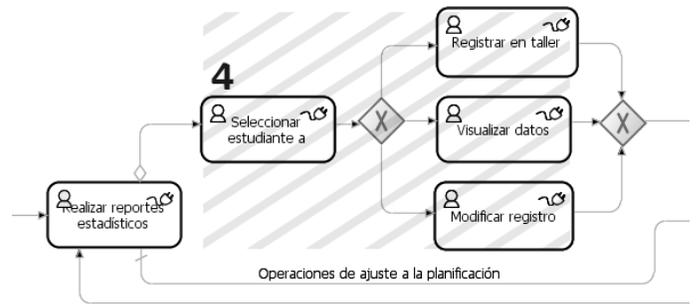


Figura 5: Segmento del Modelo Flexible Indicando con Número 4 Bifurcación

C. Fase de Construcción

Programación Lógica de Actividades del Proceso Usando Java

El motor de Flujo de trabajo de Bonita es un Java API (Application Programming Interface), que permite al usuario interactuar con el proceso. Se cuenta con distintas API's para comunicarse con el motor de Flujo de trabajo, y en este caso, se empleó la API Interface IdentityAPI para la gestión de los usuarios ya que provee métodos para añadir usuarios, manejar los grupos y roles. Cada actividad fue programada individualmente utilizando llamadas a métodos predefinidos en Java para manipular los tipos de datos utilizados, a fin de ser visualizados en los formularios web diseñados para la interacción con el usuario final (experiencia de usuario User XP), así como configurar y establecer los enlaces con la base de datos MySQL creada.

Las funciones de aplicación asociadas al proceso se exportan como un archivo *.bar (conectores, librerías, scripts o secuencias de comando) y como archivo *.war (las aplicaciones definidas como las plantillas HTML), con o sin la información de tiempo de ejecución. Estos archivos constituyen un modelo ejecutable del proceso.

Creación del Entorno de Ejecución del Proceso

El paquete Tomcat contiene tanto el motor de ejecución como el User XP, los archivos *.bar y *.war son exportados a éste paquete, obteniendo así un entorno de ejecución con Bonita.

En el caso de estudio, cuando se ejecuta la aplicación, el proceso y los recursos se empaquetan en archivos *.bar, archivos de procesos en Bonita, uno para cada pool diagramado (Coordinación--2.0.bar, Estudiante--2.0.bar, Instructor--2.0.bar), se generan *.war para cada pool y un Bonita.war que engloba a todos, un servidor Web se inicia (Apache) y el archivo Bonita.war es desplegado que encierra la aplicación web de todo el proceso.

V. CONCLUSIONES Y TRABAJOS FUTUROS

Como producto de esta investigación, el uso del enfoque de desarrollo de software basado en procesos de negocio ha sido validado, agregando aspectos de flexibilidad en tiempo de diseño que facilitan la promoción de características de calidad como la mantenibilidad, utilizando una notación estándar como BPMN y una herramienta para la gestión de procesos de negocio o BPS's de software libre (Bonita). La especificación

del proceso “Gestión del Requisito Obligatorio (RO) Apreciación Cultural de los planes de Estudio de Pregrado de la Facultad de Ciencia y Tecnología (FaCyT) de la Universidad de Carabobo”, se considera un aporte positivo a la conformación de un sistema integrado de gestión de sus procesos académicos, administrativos y de extensión de una organización como ésta. Más aun cuando es posible responder a los cambios frecuentes que ocurren, tales como continuas actualizaciones en los Planes de Estudio de Pregrado de la FaCyT, el variable número de estudiantes con opción de cursar el RO, ajustes en su planificación y desarrollo, modificaciones en los mecanismos de evaluación, entre otros.

El enfoque de flexibilidad en tiempo de diseño aplicado en esta investigación, soportado en una notación grafica como BPMN promueve la modelación visual como una de las mejores prácticas del desarrollo del software. Asimismo, el uso de una metodología como RUP enriquecida con el método BMM facilitó la construcción de la aplicación de flujo de trabajo.

REFERENCIAS

- [1] W. M. van der Aalst, *Process-Aware Information Systems: Lessons to Be Learned from Process Mining*, Transactions on Petri Nets and Other Models of Concurrency II, LNCS vol. 5460, pp. 1–26, 2009.
- [2] A. Bentellis and Z. Boufaïda, *Conceptual Method for Flexible Business Process Modeling*, Engineering and Technology, vol. 37, pp. 286–290, 2008.
- [3] M. Adams, A. ter Hofstede, N. Russell, and W. van der Aalst, *Dynamic and Context-Aware Process Adaptation*, Handbook of Research on Complex Dynamic Process Management: Techniques for Adaptability in Turbulent Environments, IGI Global, pp. 104–136, 2009.
- [4] T. Burkhart, C. Dorn, D. Werth, and P. Loos, *A Flexible Approach Towards Self-Adapting Process Recommendations*, Computing and Informatics, vol. 30, no. 1, pp. 89–111, 2011.
- [5] D. Ortega, M. Guevara y I. Durán, *Desarrollo de Aplicaciones Usando el Enfoque de Sistemas de Información Centrado en Procesos y Patrones del Paradigma Acción Comunicativa*, en las memorias del II Simposio Científico y Tecnológico en Computación (SCTC 2012), Caracas, Venezuela, pp. 196–202, 2012.
- [6] H. Schonenberg, R. Mans, N. Russell, N. Mulyar, and W. M. P. van der Aalst, *Process Flexibility: A Survey of Contemporary Approaches*, in Advances in Enterprise Engineering I, pp. 16–30, 2008.
- [7] J. M. Bhat and N. Deshmukh, *Methods for Modeling Flexibility in Business Process*, in Workshop on Business Process Modeling, pp. 1–8, 2005.
- [8] P. Kruchten, *The Rational Unified Process: An Introduction*, 2nd edition, Addison-Wesley Professional, 2000.
- [9] J. C. Montilva and J. A. Barrios, *BMM: A Business Modeling Method for Information Systems Development*, CLEI Electronic Journal, vol. 7, no. 2, pp. 1–19, 2004.
- [10] *Bonita Open Solution*, <http://www.bonitasoft.org>.
- [11] J. A. Fisteus, *Definición de un Modelo para la Verificación Formal de Procesos de Negocio*, Tesis Doctoral, Universidad Carlos III de Madrid, Departamento de Ingeniería Telemática, Leganés, España, 2005.
- [12] T. van Eijndhoven, M. E. Iacob, and M. L. Ponisio, *Achieving Business Process Flexibility with Business Rules*, in proceedings of the 12th International IEEE Enterprise Distributed Object Computing Conference (EDOC’08), Munich, Germany, pp. 95–104, September 2008.
- [13] F. Daoudi and S. Nurcan, *A Framework to Evaluate Methods Capacity to Design Flexible Process*, in proceedings of the 6th Workshop on Business Process Modeling, Development and Support (BPMDS’05), Porto, Portugal, June 2005.
- [14] K. Kumar and M. M. Narasipuram, *Defining Requirements for Business Process Flexibility*, in Workshop on Business Process Modeling, Design and Support (BPMDS’06), Proceedings of CAiSE06 Workshops, pp. 137–148, 2006.
- [15] H. Ozawa, *Designing to Attain and Maintain Business Process Flexibility*, Services Computing Conference 2009 (APSCC 2009 IEEE AsiaPacific), pp. 193–198, 2009.
- [16] S. A. White and D. Miers, *BPMN Modeling and Reference Guide*, Lighthouse Point, 2008.
- [17] E. Rolón, F. Ruiz, F. García y 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, vol. 6, pp. 1–10, 2005.

Retos para el Procesamiento Semántico de Datos Enlazados en la Nube de los Datos Abiertos Enlazados

Maria-Esther Vidal, Maribel Acosta, Ana Alvarado, Oriana Baldizán, Alexander Baranya, Simón Castillo, Giuseppe De Simone, Marlene Goncalves, Hancel González, Alexandra La Cruz, Gabriela Montoya, Guillermo Palma, Edna Ruckhaus
{mvidal, macosta, faalvarado, obaldizan, abaranya, scastillo}@ldc.usb.ve
{gdsimone, mgoncalves, hgonzalez, alacruz, gmontoya, gpalma, ruckhaus}@ldc.usb.ve

Departamento de Computación, Universidad Simón Bolívar, Caracas, Venezuela

Resumen: Este artículo describe los proyectos desarrollados por los integrantes del Grupo de la Web Semántica en la Universidad Simón Bolívar en el contexto de la Nube de los Datos Abiertos Enlazados. Los principales retos abordados en estos proyectos son los siguientes: Descomposición de consultas SPARQL y técnicas de procesamiento para escalar a federaciones con una gran número de fuentes de datos; Estrategias para explotar anotaciones semánticas de ontologías médicas para mejorar la calidad del renderizado de imágenes médicas; Técnicas de *ranking* basadas en *skyline* para de forma eficiente identificar los recursos que satisfacen consultas multi-objetivo; Técnicas de limpieza en Datos Enlazados para identificar ambigüedades entre datos enlazados y sugerir posibles inconsistencias y falta de completitud; Técnicas de Minería de grafos para descubrir patrones entre grafos anotados; Herramientas para la evaluación de la calidad y el rendimiento de las máquinas de consultas SPARQL sobre documentos SPARQL. Finalmente, se presenta un resumen de los resultados de los estudios experimentales los cuales muestran la calidad de las estrategias desarrolladas y los casos de uso donde las arquitecturas propuestas han sido aplicadas.

Palabras Clave: Web Semántica; Datos Abiertos Enlazados; Optimización y Ejecución de Consultas; Ontologías Médicas; Técnicas de Limpieza de Datos; Manejo de Imágenes Médicas; Minería de Datos.

Abstract: In this paper, we provide an overview of the projects developed by the Semantic Web group of the University Simón Bolívar in the context of the Linking Open Data Cloud, that address the following challenges: Semantic Query Decomposition and Processing techniques to scale up to federations of very large linked datasets; Strategies to exploit semantic annotations of medical ontologies to improve the quality of medical image rendering; A skyline-based ranking approach to efficiently rank resources that fulfill multi-objective queries; Linked Data Cleansing techniques to identify ambiguities among the linked data, and suggest possible inconsistencies and incompleteness; Mining techniques to discover patterns among linked annotated graphs; Benchmarking tools to evaluate the quality and performance of the existing RDF engines which are used to execute SPARQL queries against federation endpoints. Finally, we summarize experimental results that show the quality of the developed strategies and the use cases where the proposed architecture has been applied.

Keywords: Semantic Web; Linked Open Data Cloud; Query Optimization and Execution; Biomedical Ontologies; Cleaning Data; Management of Medical Images; Data Mining.

I. INTRODUCCIÓN

En la última década el número de conjuntos de datos en la nube de los datos abiertos enlazados (LOD *cloud*) ha explotado así como el número de *endpoints* de SPARQL que acceden y manejan estos conjuntos de datos. Usuarios más que nunca pueden recuperar datos que satisfacen sus requerimientos al buscar o consultar alguna de las fuentes disponibles. Esta democratización de información sienta las bases para el descubrimiento de propiedades y relaciones entre datos y enlaces que no podrían ser identificados años anteriores. Esto ha

originado problemas abiertos en las tareas tradicionales de las máquinas federadas de procesamiento de datos, *benchmarking*, visualización de imágenes, *ranking* multi-objetivo, limpieza de datos y descubrimiento de patrones.

Endpoints de SPARQL que existen actualmente para acceder y manejar conjuntos de datos enlazados, deben ser capaces de ejecutar cualquier consulta SPARQL. Sin embargo, algunos *endpoints* no son capaces de resolver consultas cuyo tiempo de ejecución o cardinalidad de la respuesta sobrepasa un cierto valor, mientras otros simplemente detienen su ejecución sin

producir respuesta alguna. Sin tener la tecnología lista para el manejo de cualquier consulta sobre los *endpoints*, existe la necesidad de desarrollar técnicas de descomposición de consultas que reescriben las consultas en subconsultas que pueden ser ejecutadas. Adicionalmente, los datos accesibles en el Web están usualmente caracterizados por: *i*) ausencia de estadísticas, *ii*) condiciones impredecibles al ejecutar las consultas remotas, y *iii*) características cambiantes.

El paradigma de optimización y luego ejecución que ha sido tradicionalmente usado para la identificación de planes de ejecución eficientes, no puede adaptarse a cambios no predecibles en los datos; así, técnicas de procesamiento de consultas adaptativas son necesarias. Aunque los *endpoints* de SPARQL y las iniciativas para impulsar el desarrollo de la nube de datos enlazados abiertos sientan las bases para el acceso de largos volúmenes de datos, existen aún aplicaciones donde es importante identificar las tuplas que mejor satisfacen una consulta o requerimiento. Basándonos en trabajos relacionados, proponemos una solución para este problema de *ranking* y desarrollamos una técnica para identificar los recursos que mejor satisfacen una consulta multi-objetivo, i.e., los recursos para los cuales no existe algún otro recurso que sea mejor en todas las condiciones del criterio multi-objetivo. Este conjunto de puntos no dominados se conoce como *skyline*.

Además, colecciones en la Nube de los Datos Abiertos Enlazados facilitan a los científicos la minería de los conjuntos enlazados para descubrir patrones o sugerir potenciales nuevas asociaciones. Para abordar el problema de descubrimiento de patrones, hemos desarrollado métricas de similitud que midan relaciones entre dos conceptos, y así, explorar las evidencias de las anotaciones, hacer predicciones, encontrar *outliers* y evaluar hipótesis. Sin embargo, para asegurar resultados confiables, datos enlazados deben cumplir altos estándares de calidad. En consecuencia, herramientas para detectar posibles ambigüedades y problemas de calidad son necesarias. Hemos implementado técnicas basadas en Redes Bayesianas y un sistema de reglas para analizar la calidad de los datos y proponer enlaces para resolver estas ambigüedades.

En este artículo describimos SEPIAS (A *Semantic Data Processing and Cleansing, analysis and visualization*) una herramienta para manejar y procesar datos enlazados en el Nube de los Datos Abiertos Enlazados. SEPIAS está compuesto por siete subsistemas que proveen técnicas para ejecutar las siguientes tareas: *i*) Descomposición y Procesamiento Semántico de Consultas, *ii*) Renderizado de Imágenes Médicas basado en Anotaciones Semánticas, *iii*) *Ranking* basado en *skyline*, *iv*) Limpieza de Datos Enlazados, *v*) Minería de Grafos Anotados, y *vi*) Herramientas de *Benchmarking*.

Este artículo está compuesto de cuatro secciones. En la Sección II, se motiva el trabajo realizado a través de un ejemplo. Luego en la Sección III, se describe SEPIAS, y se detallan cada uno de sus componentes y como los mismos interactúan. Finalmente, la Sección IV presenta nuestras conclusiones y trabajos futuros.

II. EJEMPLO MOTIVADOR

Considere la siguiente consulta expresada en SPARQL 1.0: *Seleccionar las enfermedades y los genes asociados a drogas que han sido probadas en ensayos clínicos donde el Cáncer de Próstata fue estudiado.*

```
PREFIX linkedct: <http://data.linkedct.org/resource/linkedct>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema>
PREFIX diseaseome: <http://www4.wiwiss.fu-berlin.de/diseaseome/resource/diseaseome>
(0) SELECT DISTINCT ?II ?D ?GN2
(1) WHERE {
(2) ?CT1 linkedct:condition?C1 .
(3) ?CT1 linkedct:intervention ?I .
(4) ?I linkedct:intervention_type "Drug" .
(5) ?C1 rdfs:seeAlso ?D .
(6) ?I rdfs:seeAlso ?II .
(7) ?C linkedct:condition_name "Prostate Cancer" .
(8) ?CT linkedct:intervention ?I .
(9) ?CT linkedct:condition ?C .
(10) ?D diseaseome:associatedGene ?GN2 .
(11) ?D diseaseome:possibleDrug ?II }
```

La respuesta de esta consulta está compuesta por 192 tuplas cuando los datos son recuperados de Diseaseome¹ y LinkedCT². Sin embargo, si la consulta se ejecuta contra los *endpoints* existentes de Diseaseome o LinkedCT, la respuesta es vacía. Este problema es generado por la necesidad de atravesar enlaces entre los conjuntos de datos para responder la consulta y la mayoría de los *endpoints* existentes han sido creados para consultas “ligeras” que no requieran navegar a través de varios conjuntos de datos. Otro problema que se ilustra en este ejemplo, es causado por el nivel de confianza que los usuarios tengan de los *endpoints* consultados. Por ejemplo, si los enlaces `rdfs:seeAlso` no son almacenados en estos conjuntos de datos, la respuesta podría ser incompleta, y drogas importantes podrían no haberse incluido en la respuesta. Adicionalmente, un científico podría estar interesado en analizar el grado de relación entre las enfermedades y los genes recuperados, para poder determinar posibles nuevas asociaciones, y así, por ejemplo, sugerir nuevos tratamientos. Finalmente, técnicas de *ranking* podrían ayudar a identificar las enfermedades y genes cuyas asociaciones satisfacen un cierto criterio de calidad o confianza, mientras que herramientas de *benchmarking* son requeridas para evaluar y entonar las diferentes técnicas de procesamiento que son necesarias para ejecutar estas consultas. En este artículo se describen proyectos desarrollados en el área de la Web Semántica y se abordan algunos de los retos brevemente descritos en este ejemplo.

III. LA ARQUITECTURA DE SEPIAS

La Figura 1 presenta la arquitectura de SEPIAS, un sistema para el procesamiento semántico y análisis de datos RDF los cuales son accesibles a través de *endpoints* que respetan el protocolo SPARQL. Esta arquitectura está compuesta de varios subsistemas; ANAPSID, DEFENDER, FRAGOLA, LiQuate, ANISE y Patterns in ANnotation Graphs. Usando SEPIAS el usuario puede enviar una consulta en SPARQL y obtener respuestas que corresponden a unión de los datos publicados en la federación de los *endpoints* SPARQL; las

¹ <http://www4.wiwiss.fu-berlin.de/diseaseome/sparql>

² <http://data.linkedct.org/sparql>

técnicas de descomposición, optimización y ejecución implementadas por ANAPSID y DEFENDER permiten identificar subconsultas que pueden ser ejecutadas por los *endpoints* SPARQL disponibles, planes de ejecución que minimizan el tiempo de ejecución, y ejecuciones que se adaptan a las condiciones de los *endpoints*. Las consultas SPARQL pueden ser enviadas no sólo por usuarios sino por los otros componentes de la herramienta, es decir: FRAGOLA, LiQuate, ANISE y Patterns in Annotation Graphs. FRAGOLA implementa técnicas de *ranking* para resolver consultas multi-objetivo; para recuperar los datos se comunica con DEFENDER a través de las consultas SPARQL correspondientes. LiQuate se basa en una red Bayesiana para identificar inconsistencias en datos enlazados; para recuperar los datos que permiten construir la red Bayesiana interactúa con DEFENDER a través de las consultas SPARQL. De forma similar Patterns in Annotation Graphs hace uso de DEFENDER para obtener los datos sobre los cuales se aplicarán las técnicas de minería de datos, así como, métricas de similitud para determinar relaciones entre los datos. Finalmente, ANISE hace uso de DEFENDER para recuperar la información de los términos que aparecen en las ontologías y que sirven de base para el proceso de razonamiento implementado a nivel del *segmentador*. A continuación presentamos una breve descripción de las principales propiedades que caracterizan a los subsistemas que componen a SEPIAS.

A. ANAPSID y DEFENDER Soluciones a Consultas SPARQL contra Federaciones de Endpoints

ANAPSID es una máquina de procesamiento adaptativa de datos accesibles a través de *endpoints* SPARQL [1]. ANAPSID identifica un plan de ejecución pero adapta sus operadores en base a la disponibilidad de los datos. Así, ANAPSID implementa una técnica a nivel del operador y es capaz de recuperar información desde la federación de los *endpoints* SPARQL y adaptar la ejecución de forma dinámica dependiendo de la carga de trabajo y disponibilidad de los *endpoints*. DEFENDER [12], [11] es un des-compositor de consultas sobre *endpoints* SPARQL. Primero, patrones de tripletas en la consulta se descomponen en subconsultas simples que pueden ser completamente ejecutadas por al menos un *endpoint*. Luego, las subconsultas son combinadas en un árbol tipo arbusto donde el número de *joins* se maximiza y la altura del árbol se minimiza. DEFENDER usa descripciones de los *endpoints* para seleccionar los *endpoints* SPARQL que pueden contestar cada patrón de triplete; patrones de tripletas que pueden ser ejecutadas por los mismos *endpoints* se agrupan juntas. DEFENDER se implementó sobre ANAPSID e implementa adaptividad al nivel de las fuentes, resolviendo el problema de decisión de encontrar los *endpoints* SPARQL que pueden ejecutar una descomposición dada de la consulta de acuerdo a las condiciones actuales de la federación. DEFENDER está compuesto por un *Planificador de Consultas*, un *Motor de Consultas Adaptativo* y un *Catálogo de la descripción de los endpoints*. El *Planificador de Consultas* de DEFENDER está constituido principalmente por dos componentes: el *des-compositor de consultas* y el *Optimizador de*

Consultas Basado en una Heurística. El primero divide los conjuntos de patrones de tripletas de consultas SPARQL 1.0, en subconjuntos de patrones de tripletas (sPT) que pueden ser ejecutados por un mismo *endpoint*. El des-compositor de consultas comienza creando una única subconsulta en sPT, entonces esta se mezcla con las subconsultas que comparten exactamente una variable y se repite el procedimiento hasta que se alcanza un punto fijo en el proceso de creación de subconsultas. Entonces los sPT que comparten una variable con cualquier sPT son agregados. Una vez que la consulta se vuelve a escribir en SPARQL 1.1, las técnicas de optimización basadas en heurísticas son aplicadas para generar un plan de árbol tipo arbusto, donde las hojas corresponde a subconsultas de los sPT previamente identificadas. Las técnicas de optimización no se basan en la información estadística recolectada de los *endpoints*, solamente hacen uso de la información sobre los predicados de los conjuntos de datos accesibles a través de los *endpoints*. Un algoritmo basado en una heurística ávida es implementado, éste atraviesa el espacio de planes en iteraciones y tiene como salida un plan de árbol tipo arbusto de una consulta SPARQL 1.1, en donde el número de *joins* es maximizado y la altura del árbol es minimizado. De esta forma el costo de ejecución es reducido. La Figura 2 compara el plan de ejecución generado por DEFENDER y otro generado por un sistema como FedX [17]. Cada hoja de los árboles es anotada con los *endpoints* donde las subconsultas se ejecutarán. Como puede observar, las subconsultas identificadas en el grupo de DEFENDER tienen un número más largo de patrones de tripletas. Además la altura de árbol es más pequeña en el plan producido por DEFENDER. Estas dos propiedades de los planes producidos por DEFENDER, reduce el número de datos transferidos desde los *endpoints* y puede disminuir el tiempo total de ejecución. Estudiamos el rendimiento de los planes introducidos por DEFENDER y mostramos que estos planes son competitivos con los planes generados por los motores RDF existentes. Un portal que publica nuestros resultados experimentales, puede ser visitado en <http://www.defender ldc.usb.ve>. El portal de DEFENDER presenta el comportamiento de 36 consultas SPARQL contra las colecciones de FedBench³: *Cross-Domain*, *Linked Data* y *Life Science* [16]. Estas consultas incluyen 25 consultas de FedBench y 11 consultas complejas⁴. Las consultas complejas están compuestas de entre 6 y 48 patrones de tripletas y pueden ser descompuestas en hasta 8 subconsultas. Las colecciones de FedBench fueron ejecutadas a través de los nueve (9) *endpoints* de Virtuoso⁵ con un tiempo máximo de espera de 240 segundos o 71.000 tuplas. Se consideraron diferentes federaciones donde los datos fueron particionados siguiendo diferentes criterios de fragmentación: horizontal, vertical e híbrido. Se evaluó el comportamiento de tres máquinas de ejecución de consultas: ARQ, FedX y DEFENDER. Como resultado de la evaluación es el siguiente:

Parámetros que afectan la completitud de la respuesta a las consultas. Las consultas estudiadas están confor-

³ <http://code.google.com/p/fbench>

⁴ <http://www ldc.usb.ve/~mvidal/FedBench/queries/ComplexQueries>

⁵ <http://virtuoso.openlinksw.com>

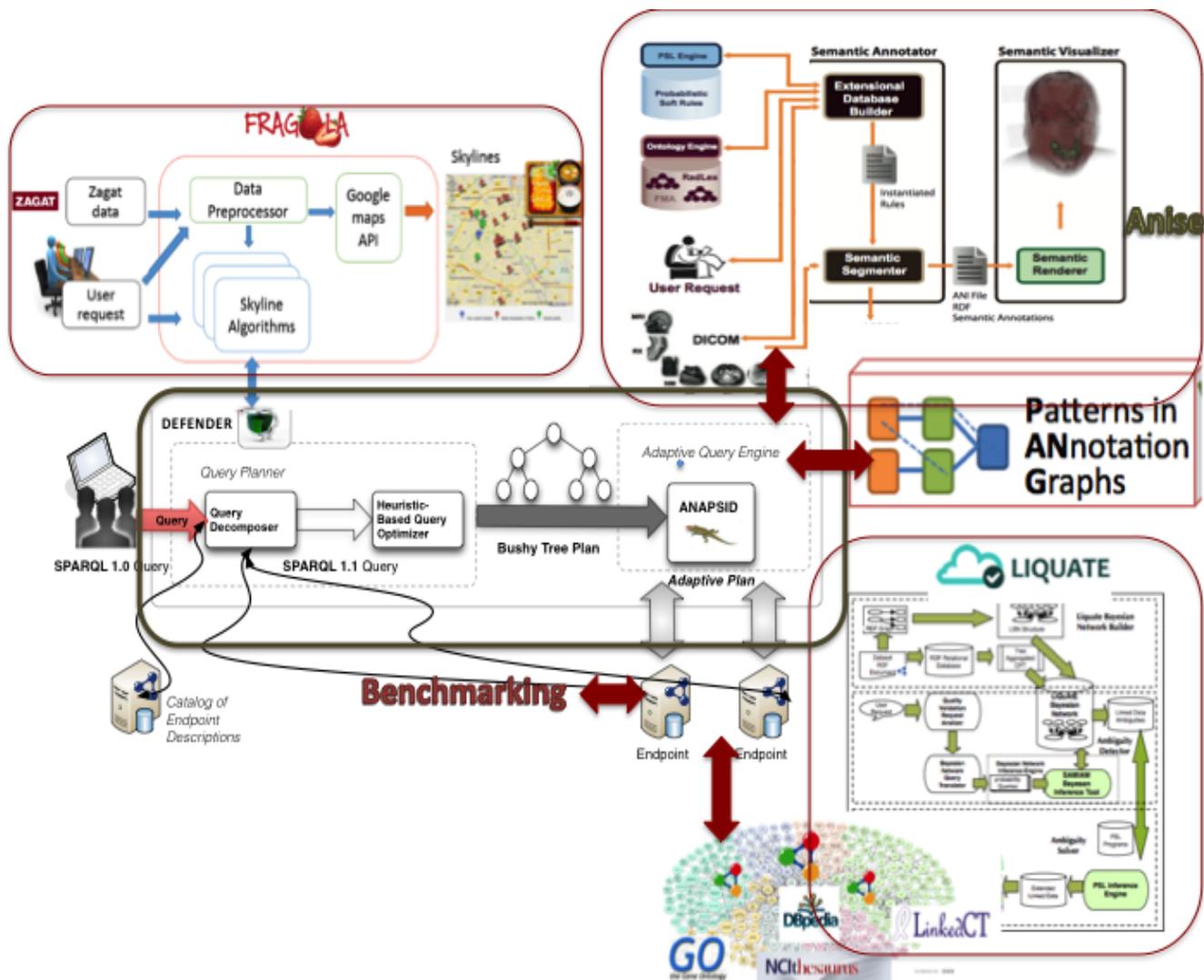
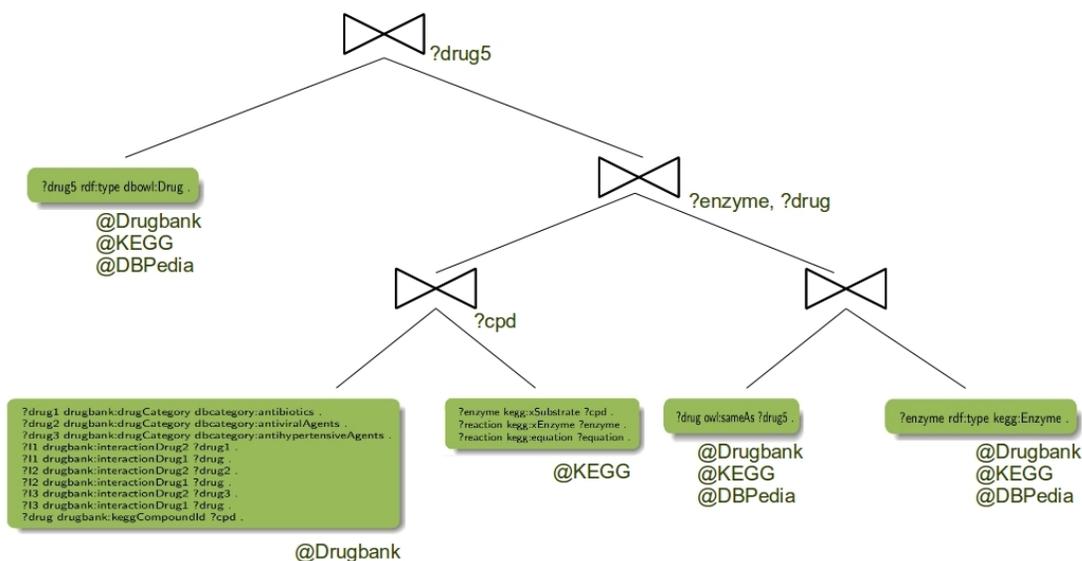


Figure 1: La Arquitectura de SEPIAS

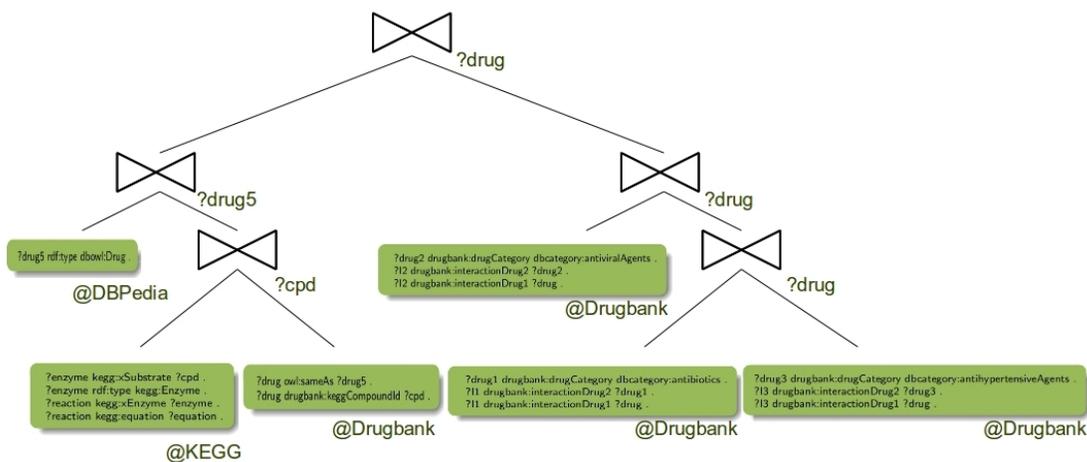
masas de predicados generales tales como `owl:sameAs` y `rdfs:seeAlso` que están presentes en la mayoría de los conjuntos de datos. Sin embargo, sólo un número muy reducido de *endpoints* que exportan tripletas con estos predicados, produce la respuesta necesaria para contestar la consulta. Esto trae como consecuencia que la máquina de ejecución debe implementar técnicas de selección de consultas para determinar cuales son los *endpoints* relevantes para evitar transferencia de datos no relevantes y garantizar completitud de las respuestas. Se pudo observar que el tiempo de ejecución de los planes producidos por FedX se puede ver afectado por este tipo de predicados, ya que esta máquina de ejecución trata de contactar todas los *endpoints*. Por otro lado, DEFENDER implementa distintas heurísticas para determinar cuales son los *endpoints* relevantes; las estrategias implementadas permiten reducir el tiempo de ejecución pero en algunos casos comprometen la completitud de la respuesta de la consulta. En conclusión pudimos observar que el tipo de predicados en los patrones de tripletas de las consultas pueden afectar el

comportamiento de las máquinas de ejecución en ambientes de federaciones de *endpoints*.

Efectos del tipo de consulta en el tiempo de ejecución y en la completitud de la consulta. Se estudiaron planes de ejecución optimales para las diferentes consultas, es decir, planes lineales izquierdos y tipo arbusto. Estos planes fueron evaluados en ARQ y DEFENDER. Se pudo observar que los planes tipo arbusto reducen el tiempo de ejecución en al menos un orden de magnitud en ambas máquinas. También el estudio sugiere que DEFENDER puede ejecutar de manera más eficiente consultas con un gran número de patrones de tripletas; cabe destacar que FedX y ARQ no son capaces de producir respuesta en al menos en 30 minutos para las consultas complejas estudiadas. Finalmente, el tiempo requerido para ejecutar las subconsultas identificadas por DEFENDER es mucho menor que el requerido por FedX y ARQ. En base a los resultados observados se puede concluir que el tipo de consulta si afecta el comportamiento de los máquinas de ejecución.



(a) Planes Basados en Grupos Exclusivos (Sistemas como FedX)



(b) Planes Producidos por DEFENDER

Figure 2: Comparación de Planes Producidos por Sistemas como FedX y DEFENDER

B. Fragola: Una Propuesta para Identificar Los Mejores Lugares Gastronómicos

FRAGOLA es un sistema de clasificación jerárquica que implementa el algoritmo *skyline* FOPA (*Final Object Pruning Algorithm*) [2] que es capaz de producir de manera eficiente todo el conjunto de datos de los puntos de *skyline* y escala hasta grandes conjuntos de datos. FOPA se basa en tablas particionadas verticalmente ordenadas para almacenar datos RDF recuperados de la federación de *endpoints* SPARQL, información sobre los valores observados hasta ahora e índices de la tablas particionadas verticalmente para podar el espacio de los puntos dominantes. FOPA es apto para identificar el *skyline* para conjuntos de datos grandes en menos tiempo que los enfoques del estado del arte. Como una prueba del concepto, desarrollamos FRAGOLA (*Fabulous RAnking of Gastronomy LocAtions*), una herramienta que implementa FOPA y clasifica jerárquicamente lugares gastronómicos basado en criterios multi-objetivos.

El sistema FRAGOLA fue desarrollado sobre el conjunto de datos multidimensional de restaurantes en París que es proporcionado por Zagat⁶. La ontología del restaurante⁷ es usada para describir cada restaurante, mientras que Geonames⁸ es usado para describir la localización geo-espacial. La respuesta a una consulta multi-objetivo es un conjunto de restaurantes que no son comparables, que componen el *skyline* con respecto a los atributos y directivas consideradas en la consulta. FRAGOLA está compuesto de los siguientes componentes: i) un *Pre-procesador de Datos*, ii) un *Motor de Skyline* y iii) el *API de Google Maps*. El pre-procesador de datos transforma los datos proporcionados por Zagat en la escala [0, 1] y calcula la distancia entre la posición geo-espacial actual de un usuario y la de cada restaurante. Un motor de *skyline* implementa tres algoritmos de *skyline*: IDSA, desarrollado por Balke et al [4], RSJFH, introducido por Chen et al [7] y FOPA. El API de

⁶ <http://www.zagat.com/paris>

⁷ <http://schema.org/Restaurant>

⁸ <http://www.geonames.org>

Google Maps es usado para visualizar el conjunto *skyline* de restaurantes en el mapa. La Tabla I reporta el número promedio de: lecturas, *joins*, comparaciones y podas realizados por cada algoritmo; los mejores resultados se resaltan en **negritas**. Se observa que FOPA puede reducir el número de comparaciones hasta en dos órdenes de magnitud, mientras que el número de lecturas y *joins* es reducido en un orden de magnitud; los resultados observados sugieren que la combinación de ordenamiento de los datos y técnicas de poda pueden reducir el tipo de ejecución de consultas multi-objetivo. El demo está publicado en <http://fragola ldc.usb.ve>.

Table I: Rendimiento de RSJFH, IDSA y FOPA

Algoritmos	Lecturas	Joins	Comparaciones	Podas
RSJFH	119	119	139	107
IDSA	49	40	5417	0
FOPA	28	27	70	16

C. Limpiando Datos Enlazados con LiQuate

LiQuate es una herramienta capaz de identificar ambigüedades entre la datos enlazados de la federación de *endpoints* SPARQL y sugiere posibles inconsistencias e incompletitudes. LiQuate también implementa un enfoque de doble pliegue que combina Redes Bayesianas y sistemas basados en reglas para analizar la calidad de los datos y proponer nuevos enlaces que resuelven las ambigüedades detectadas. LiQuate ha sido construido sobre los conjuntos de datos enlazados de *Life Science* que mantiene datos relacionados con intervenciones, condiciones, drogas, enfermedades y las relaciones entre ellos. LiQuate recibe una *solicitud de validación de calidad* la cual es expresada como una o más consultas de evidencias contra la Red Bayesiana. La respuesta de una *solicitud de validación de calidad* es un número en el rango [0.0 : 1.0] que indica la probabilidad de que los datos y enlaces sean inconsistentes o incompletos. Actualmente pueden ser expresadas tres tipos de solicitudes de validación de calidad: *i*) probabilidad de que una etiqueta o nombre de un tipo o recurso es redundante, *ii*) probabilidad de enlaces incompletos en un conjunto de recursos dado y *iii*) probabilidad de enlaces inconsistentes. LiQuate está compuesto por tres componentes: el *Constructor de Redes Bayesianas de LiQuate*, el *Detector de Ambigüedad* y el *Resolvedor de Ambigüedad*. El demo de LiQuate está publicado en <http://liquate ldc.usb.ve>.

D. Descubriendo Patrones en Datos Enlazados

Patterns in ANnotation Graphs [3] es una herramienta que permite a científicos identificar patrones in conjuntos de datos de grafos anotados. Éste provee diferentes métricas de similitud para computar distancias o el grado de relación entre dos conceptos científicos. Los grafos anotados son obtenidos de la federación de *endpoints* SPARQL, haciendo uso de las técnicas de procesamiento semántico de consultas implementadas por DEFENDER y ANAPSID. Un primer paso para descubrir patrones complejos requiere determinar el grado de relación (o similitud) de un par de conceptos, basado en sus anotaciones con respecto a una o

más ontologías. Un ejemplo es identificar el grado de relación o similitud entre pares (droga, droga), basado en la anotación de evidencia de enfermedades (condiciones) en una ontología dada. Esto puede guiar a descubrimientos de nuevos objetivos para drogas existentes, o se puede predecir potenciales efectos colaterales de drogas existentes. Nosotros abordamos el reto del análisis de datos enlazados a gran escala de los conjuntos de datos de grafos anotados, usando el conocimiento semántico de las ontologías. Para explicar esto, considere dos drogas Brentuximab vedotin y Catumaxomab. La Figura 3 representa un subgrafo de un grafo de anotación; intervenciones están en verde, condiciones están en rectángulos rosados, y términos de la ontología NCI *Thesaurus* están encerrados en óvalos rojos. Cada camino entre un par de condiciones, ej., Carcinoma y Anaplastic Large Cell Lymphoma a través de NCI *Thesaurus* es identificado usando círculos rojos los cuales representan términos de la ontología NCI *Thesaurus*. El número de círculos rojos representa la longitud del camino. Para simplificar la Figura, sólo ilustramos los caminos desde el término Carcinoma.

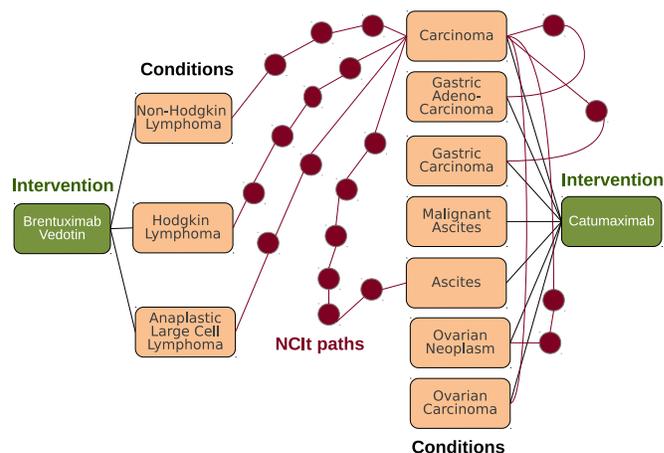


Figure 3: Subgrafo de Anotaciones de las Drogas Brentuximab Vedotin y Catumaxomab

Nuestro objetivo es identificar patrones de conectividad que puedan ser utilizados para proporcionar un resumen completo de la relación de los conceptos conectados. Esta investigación ha sido reportada en [6], [13], [14]. Definimos dos tipos de métricas de similitud: *i*) métricas de similitud topológica que mide el grado de relación en términos de la cercanía de los términos de vocabulario controlado en una taxonomía o una ontología dada y *ii*) métricas de similitud basadas en anotaciones que miden el grado de relación entre dos entidades anotadas en términos de la similitud de sus anotaciones.

En primer lugar definimos una métrica de distancia taxonómica llamada d_{tax} [6]. La intuición detrás de la métrica d_{tax} es capturar la distancia taxonómica entre dos vértices con respecto a la profundidad del ancestro común de estos dos vértices. Además, d_{tax} trata de asignar los más bajos valores de distancia taxonómica a pares de vértices que están (1) a una profundidad más grande en la ontología (2) están más cercanos a su ancestro común más bajo. Un valor cercano a 1.0 indica que ambos vértices son generales o que que el ancestro

Table II: Similitud Conjunto de Datos 1: $(1 - d_{tax})$ para SNOMED, MeSH, y NCI

Términos Médicos	Phy	Cod	SNOMED	MeSH	NCIt
Renal Insufficiency - Kidney Failure	4.00	4.00	1.00	1.00	1.00
Heart - Myocardium	3.30	3.00	0.77	0.80	0.20
Stroke - Infarction	3.00	2.80	0.31	0.80	0.87
Abortion - Miscarriage	3.00	3.30	0.89	0.00	0.92
Delusions - Schizophrenia	3.00	2.20	0.00	0.00	0.80
Congestive heart failure - Pulmonary edema	3.00	1.40	0.50	0.00	0.59
Metastasis - Adenocarcinoma	2.70	1.80	0.83	0.25	0.00
Calcification Stenosis	2.70	2.00	0.55	0.00	0.40
Diarrhea - Stomach cramps	2.30	1.30	0.29	0.75	0.42
Mitral Stenosis - Atrial Fibrillation	2.30	1.30	0.63	0.50	0.53
Chronic obstructive pulmonary disease - Lung infiltrates	2.30	1.90	0.70	-	0.13
Rheumatoid Arthritis - Lupus	2.00	1.00	0.50	0.00	0.86
Brain tumor - Intracranial hemorrhage	2.00	1.30	0.63	0.63	0.17
Carpal Tunnel Syndrome - Osteoarthritis	2.00	1.00	0.33	0.00	0.33
Diabetes Mellitus - Hypertension	2.00	1.00	0.64	0.00	0.17
Acne - Syringe	2.00	1.00	0.00	0.00	0.00
Antibiotic - Allergy	1.70	1.00	0.00	0.00	0.00
Cortisone - Total knee replacement	1.70	1.00	0.00	0.00	0.00
Pulmonary Embolism - Myocardial Infarction	1.70	1.20	0.36	0.29	0.63
Pulmonary Fibrosis - Lung Cancer	1.70	1.40	0.75	0.67	0.60
Cholangiocarcinoma - Colonoscopy	1.30	1.00	0.00	0.00	0.00
Lymphoid hyperplasia - Laryngeal cancer	1.30	1.00	0.43	0.00	0.36
Multiple Sclerosis - Psychosis	1.00	1.00	0.44	0.00	0.33
Appendicitis - Osteoporosis	1.00	1.00	0.31	0.00	0.50
Rectal polyp - Aorta	1.00	1.00	0.00	-	0.00
Xerostomia - Liver Cirrhosis, Alcoholic	1.00	1.00	0.00	0.00	0.14
Peptic Ulcer - Myopia	1.00	1.00	0.23	0.00	0.15
Depression- Cellulitis	1.00	1.00	0.00	0.00	0.31
Varicose vein - Entire knee meniscus	1.00	1.00	0.13	-	0.00
Hyperlipidemia - Metastasis	1.00	1.00	0.33	0.00	0.00

común más bajo es cercano a la raíz de la ontología. En consecuencia, $(1 - d_{tax})$ será usado como métrica de similitud ontológica o grado de relación entre dos nodos. En segundo término, proponemos una nueva métrica de similitud llamada *AnnSim*, que mide el grado de relación entre dos entidades en términos de la similitud o grado de relación del conjunto de sus anotaciones. Modelamos *AnnSim* como un apareamiento 1-a-1 de máximo peso de un grafo bipartito. La computación de *AnnSim* primero requiere construir a grafo bipartito completo, computando todos los pares de similitudes entre los términos y luego entonces determinando el apareamiento 1-a-1 de máximo peso del grafo bipartito. Realizamos un estudio experimental para evaluar la calidad de las métricas definidas; estudiamos la capacidad de predicción de nuestras métricas en varios conjuntos de datos.

Conjunto de datos 1: 30 pares de enfermedades del *Mayo Clinic Benchmark*; cada par se codifica por su similitud desde 1.0 (menos similar) hasta 4.0 (más similar). La codificación fue llevada a cabo por 3 médicos (**Phy**) y 10 codificadores médicos de la Clínica Mayo (**Cod**) [10], [15]. Las enfermedades fueron anotadas con NCI *Thesaurus* versión 12.05d⁹. El conjunto de datos 1 es usado para comparar $(1 - d_{tax})$ usando SNOMED¹⁰, MeSH¹¹, and the NCI *Thesaurus*.

Conjunto de datos 2: 12 drogas anticancer en la intersección de anticuerpos monoclonales y agentes antineoplásicos: Alemtuzumab, Bevacizumab, Brentuximab vedotin, Cetuximab, Catumaxomab, Edrecolomab, Gemtuzumab, Ipilimumab,

Ofatumumab, Panitumumab, Rituximab, and Trastuzumab. Las drogas fueron asociadas con condiciones y enfermedades en ensayos clínicos de LinkedCT de septiembre 2011 y cada enfermedad fue enlazada a su correspondiente término en el NCI *Thesaurus* versión 12.05d.

Primero, anotamos las 30 enfermedades del conjunto de datos 1 con sus términos correspondientes de SNOMED, MeSH y NCI *Thesaurus*. Las puntuaciones determinadas por $(1 - d_{tax})$ fueron comparadas con la evaluación realizada por los médicos y codificadores. Los resultados revelan que d_{tax} es exitoso en la computación de valores altos de similitud para los pares que también son clasificados con alta puntuación por los médicos y codificadores. La Tabla II ilustra los resultados; celdas Vacías (-) representan términos que no aparecen en la ontología; valores resaltados en **negrillas** muestran una correlación relevante dada por el médico, el codificador y la métrica correspondiente. Además, condujimos una evaluación extensiva de patrones de conectividad en el conjunto de datos 2. Observamos que *AnnSim* consistentemente asigna altos valores de similitud a drogas que son usadas para tratar enfermedades similares. Los detalles del conjunto de datos 2 como sus anotaciones y los valores de similitud entre los pares que se obtienen con *AnnSim* pueden encontrados en <http://pang.umiacs.umd.edu/AEDdemo.html>.

E. ANISE: Una Herramienta para el Procesamiento de Imágenes Médicas

ANISE [5] es una herramienta que permite anotar semánticamente imágenes médicas, con el fin de permitir la visualización del tejido de interés con mejor precisión y calidad. ANISE se basa en la utilización del conocimiento

⁹ <http://ncit.nci.nih.gov>

¹⁰ http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html

¹¹ <http://www.nlm.nih.gov/mesh>

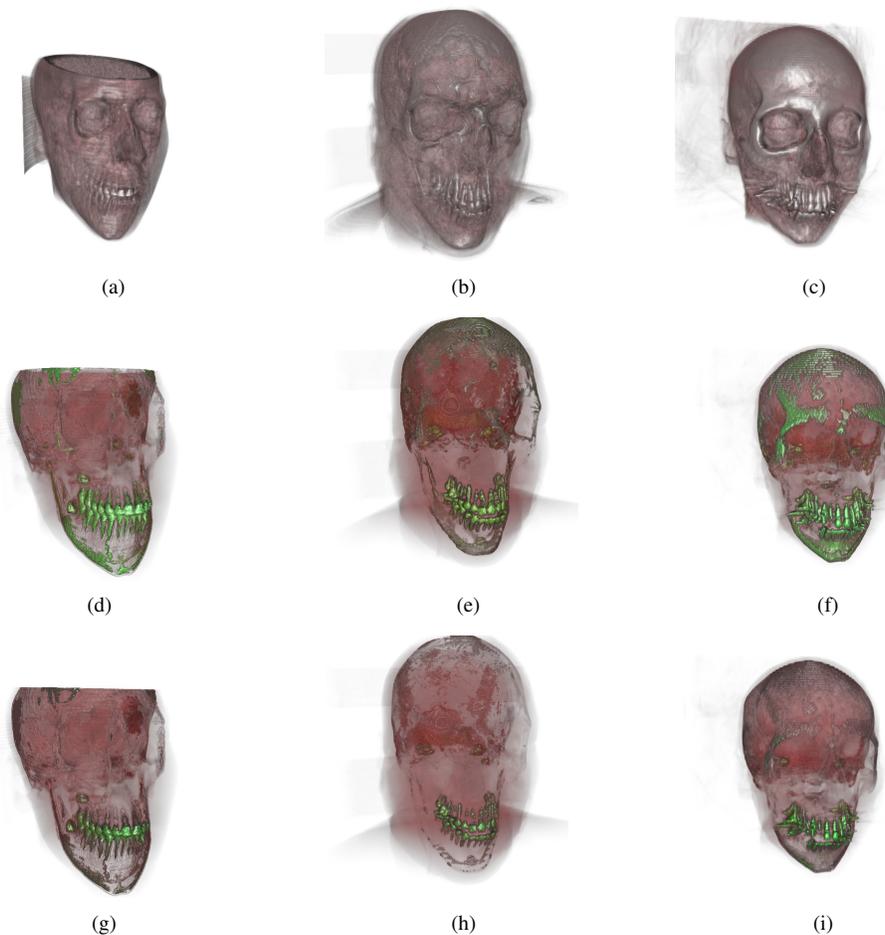


Figure 4: Resultados de Ejecutar las Técnicas Propuestas en Tres Volúmenes de Imágenes: (a) skewed_head (b) visible_human and (c) ct_head. Resultados del renderizado de tejidos usando el sistema de reglas aplicadas a: applied to (a) (b) (c) en los volúmenes mostrados en (d) (e) (f). Resultados de aplicar las reglas de segmentación se muestran en: (g) (h) (i)

expresado en las ontologías médicas FMA¹² y RadLex¹³ y un sistema basado en reglas definidos en PSL [9]. Como resultado se obtiene los datos volumétricos particionados en un sub-volumen, el cual contiene el tejido u órgano de interés anotado, utilizando los términos correspondientes de las ontologías FMA y RadLex. La estrategia implementada en ANISE está compuesta por tres fases: *i*) segmentación semántica *ii*) descripción de los recursos, y *iii*) renderizado semántico del volumen. ANISE recibe como entrada un archivo DICOM, datos del usuario y un conjunto de ontologías. El archivo DICOM corresponde a la imagen médica, y está compuesto por metadatos que describen las características de la imagen, así como la región sub-volumétrica del área que enmarca el tejido de interés y un punto semilla. Los datos del usuario indican el tejido que se quiere resaltar de la imagen

y las propiedades ópticas a utilizar para su visualización. Suponiendo que el área que contiene al tejido de interés y el punto semilla pueden ser definidos manualmente por un experto radiólogo o por una máquina de aprendizaje como la propuesta por Criminisi et al. [8]. Nosotros estudiamos la calidad de ANISE en tres imágenes médicas de tomografías computarizada del cráneo. Estas imágenes son comúnmente utilizadas en el área de computación gráfica para estudiar la calidad de las tareas segmentación y *rendering*¹⁴. El estudio tiene como objetivo mostrar la calidad de la visualización cuando los dientes son visualizados en un color particular. Dado que la densidad de los dientes es la misma que la de otros tejidos, la visualización puede ser no precisa y nuestro objetivo es estudiar el impacto de la semántica codificada en las ontologías médicas FMA y RadLex en la calidad de la

¹² <http://sig.biostr.washington.edu/projects/fm/AboutFM.html>

¹³ <https://www.rsna.org/RadLex.aspx>

¹⁴ <http://www9.informatik.uni-erlangen.de/External/vollib>

visualización. La idea es que ANISE hace uso de un sistema de razonamiento sobre lógica imprecisa y el sistema manejador de ontologías de Jena, para determinar el órgano o tejido encerrado por en las diferentes regiones que conforman la imagen. La Figura 4 ilustra el resultado del proceso de visualización tradicional y el implementado en ANISE. Figuras 4 (a), (b) and (c) muestra las imágenes estudiadas, mientras que las Figuras 4 (d), (e) and (f) presentan las imágenes donde los dientes se visualizan de verde haciendo uso del proceso tradicional de visualización. Finalmente, las Figuras 4 (g), (h) and (i) ilustran las imágenes resultantes de aplicar las técnicas propuestas. Como puede observarse, ANISE permite realizar una visualización más precisa de los tejidos que pertenecen a los dientes.

IV. CONCLUSIONES Y TRABAJOS FUTUROS

Presentamos SEPIAS, una arquitectura que integra herramientas para consultar, curar, analizar y visualizar datos en la Nube de los Datos Abiertos Enlazados. Abordamos los retos de escalar conjuntos de datos enlazados muy grandes, gestionar, clasificar y procesar federaciones de *endpoints* de SPARQL. Además, problemas como la limpieza de los datos enlazados, descubrir patrones de conectividad y visualizar imágenes médicas han sido considerados. En todos los casos la semántica codificada en los conjuntos de datos fueron usadas para mejorar la calidad de los técnicas implementadas. Reportamos estudios experimentales que revelan el rendimiento y la calidad de los enfoques propuestos. Resultados empíricos iniciales corroboran nuestra hipótesis que la semántica codificada en las ontologías y en los conjuntos de datos son importantes en cada una de las tareas que intentamos resolver. Los efectos de la semántica no sólo pueden ser observados en el tiempo de ejecución de una consulta, sino también en la calidad de la salida producida por cada herramienta.

En el futuro planificamos extender las técnicas de consultas y clasificación propuestas para tratar datos dinámicos, ej., datos de transmisión producidos por redes de sensores. Además, tenemos previsto definir otras métricas de similitud para identificar la relación en los datos enlazados que están anotados con términos de diferentes ontologías. Finalmente, ANISE será extendido con un nueva estrategia de renderizado y visualización. Asimismo, serán agregadas nuevas reglas de segmentación que describen propiedades particulares de nuevos tejidos. En consecuencia, ANISE será capaz de segmentar nuevos tejidos, por ejemplo, vasos sanguíneos.

AGRADECIMIENTO

Esta investigación ha sido parcialmente apoyada por DID-USB.

REFERENCES

- [1] M. Acosta, M.-E. Vidal, T. Lampo, J. Castillo, and E. Ruckhaus, *ANAPSID: AN Adaptive query Processing engine for SPARQL endpoints*, in proceedings of the International Semantic Web Conference (ISWC), Bonn, Germany, 2011.
- [2] A. Alvarado, O. Baldizan, M. Goncalves, and M.-E. Vidal, *FOPA: A Final Object Pruning Algorithm to Efficiently Produce Skyline Points*, in proceedings of the 24th International Conference of Database and Expert Systems Applications (DEXA), Prague, Czech Republic, 2013.
- [3] P. Anderson, A. Thor, J. Benik, L. Raschid, and M.-E. Vidal, *PAnG: Finding Patterns in Annotation Graphs*, in proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2012, Scottsdale, USA, 2012.
- [4] W.-T. Balke, U. Guntzer, and J. X. Zheng, *Efficient Distributed Skylining for Web Information Systems*, in proceedings of the International Conference on Extending Database Technology (EDBT), Heraklion, Greece, 2004.
- [5] A. Baranya, L. Landaeta, A. L. Cruz, and M.-E. Vidal, *A Workflow for Improving Medical Visualization of Semantically Annotated CT-Images*, in proceedings of the Joint Workshop on Semantic Technologies Applied to Biomedical Informatics and Individualized Medicine (SATBI+SWIM) collocated with the International Semantic Web Conference (ISWC), Boston, USA, 2012.
- [6] J. Benik, C. Chang, L. Raschid, M. E. Vidal, G. Palma, and A. Thor, *Finding Cross Genome Patterns in Annotation Graphs*, in proceedings of Data Integration in the Life Sciences (DILS), College Park, USA, 2012.
- [7] L. Chen, S. Gao, and K. Anyanwu, *Efficiently Evaluating Skyline Queries on RDF Databases*, in proceedings of the Extended Semantic Web Conference (ESWC) 2011, Heraklion, Greece, 2011.
- [8] A. Criminisi, J. Shotton, and E. Konukoglu, *Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning*, Foundations and Trends in Computer Graphics and Vision, vol. 7, no. 2-3, 2012.
- [9] A. Kimmig, S. H. Bach, M. Broecheler, B. Huang, and L. Getoor, *A Short Introduction to Probabilistic Soft Logic*, in NIPS Workshop on Probabilistic Programming: Foundations and Applications, 2012.
- [10] B. McInnes, T. Pedersen, and S. Pakhomov, *UMLS-Interface and UMLS-Similarity: Open Source Software for Measuring Paths and Semantic Similarity*, in proceedings of the AMIA Symposium, 2009.
- [11] G. Montoya, M.-E. Vidal, and M. Acosta, *A Heuristic-Based Approach for Planning Federated SPARQL Queries*, in proceedings of the Third International Workshop on Consuming Linked Data (COLD) collocated with the International Semantic Web Conference (ISWC), Boston, USA, 2012.
- [12] G. Montoya, M.-E. Vidal, and M. Acosta, *DEFENDER: a DEcomposer For quERies agaiNst feDERations of endpoints*, in proceedings of Workshops and Demo Papers at the Extended Semantic Web Conference (ESWC), Heraklion, Greece, 2012.
- [13] G. Palma, M.-E. Vidal, L. Raschid, and A. Thor, *Exploiting Semantics from Ontologies and Shared Annotations to Partition Linked Data*, in proceedings of Data Integration in the Life Sciences (DILS), Lisbon, Portugal, 2014.
- [14] G. Palma, M.-E. Vidal, E. Haag, L. Raschid, and A. Thor, *Measuring Relatedness between Scientific Entities in Annotation Datasets*, in ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics. ACM-BCB, Washington, USA, 2013.
- [15] T. Pedersen, S. Pakhomov, S. Patwardhan, and C. Chute, *Measures of Semantic Similarity and Relatedness in the Biomedical Domain*, Journal of Biomedical Informatics, vol. 40, no. 3, 2007.
- [16] M. Schmidt, O. Görlitz, P. Haase, G. Ladwig, A. Schwarte, and T. Tran, *FedBench: A Benchmark Suite for Federated Semantic Data Query*, in proceedings of the International Semantic Web Conference (ISWC), Bonn, Germany, 2011.
- [17] A. Schwarte, P. Haase, K. Hose, R. Schenkel, and M. Schmidt, *FedX: Optimization Techniques for Federated Query Processing on Linked Data*, in proceedings of the International Semantic Web Conference (ISWC), Bonn, Germany, 2011.

OCTOPUS: Aplicación de SQLf para Selección de Fuentes de Datos

Marlene Goncalves, Leonid Tineo
mgoncalves@usb.ve, leonid@usb.ve

Departamento de Computación y Tecnología de la Información, Universidad Simón Bolívar, Caracas, Venezuela

Resumen: La Web es un repositorio de datos muy grande compuesto de muchos documentos que se almacenan por varias fuentes de datos. Los motores de búsqueda Web permiten la recuperación basada en palabras clave. Sin embargo, los documentos publicados pueden ser incompletos, obsoletos o enormes. Por lo tanto, la búsqueda también podría incluir criterios de calidad como completitud, frecuencia de actualización, granularidad y cuán recientes ellos son. Presentamos Octopus, una herramienta para la selección de fuentes de datos y documentos en términos de preferencias del usuario que involucran parámetros de calidad almacenados en un catálogo relacional. La recuperación se haría por medio de SQLf, un lenguaje de consultas basado en lógica difusa. Octopus ofrece un asistente para facilitar la expresión de requerimientos de usuarios finales, así como los términos difusos involucrados, los cuales también pueden ser especificados en XML, y en el catálogo fuente de datos.

Palabras Clave: Web; SQLf; Selección de Fuentes de Datos.

Abstract: The Web is a very large data repository composed of many documents that are stored by several data sources. Web search engines allow retrieval based on keywords. Nevertheless, published documents may be incomplete, obsolete or huge. Therefore, the search criteria could also include quality and completeness, update frequency, granularity and how they are recent. We present Octopus, a tool for the selection of data sources and documents in terms of user preferences that involve quality parameters stored in a relational catalog. The recovery would SQLf through a query language based on fuzzy logic. Octopus provides a wizard to facilitate the expression of end-user requirements and fuzzy terms involved, which can also be specified in XML, and the data source catalog.

Keywords: Web; SQLf; Data Source Selection.

I. INTRODUCCIÓN

Las bases de datos tradicionales son repositorios de datos que almacenan información en tablas que implementan relaciones basadas en la teoría de conjuntos. En este sentido, las búsquedas de información retornan conjuntos de resultados donde los registros de datos del repositorio son simplemente elementos que pertenecen o no pertenecen a ellos, según los criterios de búsqueda. Según diversos autores como [1][2], y otros que tienen contribuciones en el libro de Galindo [3], esta forma de ver las cosas está distante de la manera natural como el humano percibe y clasifica la información que, usualmente, tiende a ser más flexible involucrando términos que por su naturaleza no son expresables con conjuntos clásicos.

Es aquí donde entra la teoría de las bases de datos difusas, la cual se fundamenta en la teoría de conjuntos difusos para dar soporte a requerimientos vagos e información imprecisa. En la actualidad no existe un lenguaje estándar para bases de datos difusas, tal como lo es SQL para las bases de datos relacionales clásicas. Distintos investigadores han hecho sus propias

propuestas. Entre ellas la más completa por su diversidad de expresiones de consultas difusas es SQLf, el cual ha sido reportado, entre otros trabajos previos [1][4][5].

Según [6], las búsquedas de fuentes de datos en Internet podrían ser muy favorecidas por el uso de bases de datos difusas, pues en general el usuario no tiene precisión sobre sus requerimientos sino que éstos suelen ser vagos (o flexibles). Para ello hemos concebido el sistema Octopus el cual ha sido desarrollado con SQLf usando la implementación denominada SQLfi [7][8]. En Octopus, las búsquedas se hacen de forma flexible según criterios del usuario. El sistema se encarga de buscar documentos, fuentes de datos contenedoras de estos documentos y enlaces que los apunten. Algunos trabajos anteriores [9][10] se han hecho en este sentido, pero sin involucrar consultas difusas, donde se sentaron las bases para este tipo de sistemas. Los usuarios tienen opciones de solicitar que estos elementos cumplan con condiciones difusas como “antiguo”, “reciente”, “muy actual”, condiciones en el tamaño de los archivos, entre otras, y que concuerden con sus palabras claves. Del mismo modo, los usuarios registrados en Octopus

tienen la opción de definir predicados de naturaleza difusa como “reciente”, “antiguo”, “muy”, y otros más; según sus propios criterios, de otro modo, se emplearán los que haya definido el administrador anteriormente.

Este artículo lo hemos estructurado en secciones como sigue. La Sección II, explica la concepción de Octopus; la Sección III, describe las funcionalidades de Octopus; la Sección IV, presenta los términos difusos de SQLf en Octopus; y la Sección V reseña las conclusiones y trabajos futuros.

II. CONCEPCIÓN DE OCTOPUS

Octopus es un sistema que funciona bajo la misma filosofía de buscadores existentes. Al igual que estos, un usuario ingresa al sistema, indica una serie de palabras claves, señala si desea buscar un documento o una fuente de datos que cumplan con ciertas condiciones, y el sistema se encarga de consultar la base de datos para desplegar en pantalla las opciones encontradas que cumplan con lo indicado por el usuario. Los usuarios cuentan con la opción de realizar búsquedas avanzadas, en las cuales el sistema pide muchas especificaciones sobre las cuales hacer las búsquedas, de manera que el usuario pueda recibir una respuesta más específica de lo que busca.

Octopus tiene un aspecto que lo diferencia de buscadores tradicionales: lenguaje de consultas difusas. Mediante esta herramienta, los usuarios pueden especificar búsquedas de manera más flexible, indicándole al sistema parámetros difusos que permitan obtener resultados que se ajusten “un poco más” o “un poco menos” a lo solicitado por el usuario; resultados que encajen aproximadamente (en forma gradual) con los parámetros especificados, en vez de resultados que encajen perfectamente (de forma rígida) con tales especificaciones.

Al seleccionar la opción de búsqueda avanzada, el sistema le permite al usuario especificar si desea buscar documentos, fuentes de datos y/o enlaces, cuántos de estos y en qué orden, y especificar qué información correspondiente a estos desea que le sean mostrados en pantalla (es decir, especificar las tablas que desea consultar y los atributos que desea le sean mostrados y de qué modo agruparlos). Adicionalmente, sobre los distintos atributos de las distintas tablas, el usuario tiene la opción de hacer especificaciones difusas de lo que desea conseguir, es decir, hacer uso de los predicados difusos como lo son “reciente”, “muy antiguo”. Seguidamente, con los resultados de la búsqueda, el usuario puede solicitar al sistema efectuar operaciones de conjunto sobre los elementos obtenidos (tales como la unión y la intersección) a fin de refinar su búsqueda y obtener resultados más ajustados a sus preferencias.

Si el usuario no está registrado en el sistema, o no ha iniciado sesión, las búsquedas se efectúan en base a parámetros difusos predefinidos por el administrador del sistema. De lo contrario, un usuario registrado puede hacer uso de los parámetros difusos que él mismo haya especificado anteriormente, así como crear nuevos, y editar o eliminar los que ya haya creado.

A. Descripción de los Datos

Las consultas se efectúan sobre un catálogo que registra de fuentes de datos. Éstas son caracterizadas por — Nombre: nombre de la fuente de datos, — URL: ubicación en Internet de la fuente de datos, — País: nación donde se encuentra la fuente de datos, — Ciudad: ciudad donde se encuentra la fuente de datos, — Recentness: cuán reciente fue creada la fuente de

datos, — Completitud: cuán completa es la información contenida en la fuente de datos, — Frecuencia: con cuánta frecuencia se actualiza la fuente de datos, — Granularidad: cuán dispersos se encuentran los documentos en las fuentes de datos, — Unidad de Granularidad: orden de magnitud de almacenamiento (bytes, k-bytes, g-bytes, etc.) en que se encuentran la mayoría de los documentos contenidos en la fuente de datos.

Los documentos son descritos mediante: — Título: nombre del documento, — URL: ubicación del documento, — Autor: nombre de persona que lo creó, — Idioma: lengua en que le documento está escrito, — Tipo de archivo: Formato en que se encuentra almacenado el documento, — Fecha de creación: Fecha en que fue creado el documento, — Tamaño: Espacio que ocupa el documento en memoria, — Número de páginas: Número de páginas que ocupa el texto del documento, — Enlaces: Los atributos de un enlace dependen de si dicho enlace lleva hacia un documento o hacia una fuente de datos, y dependerán de ello.

Existen formas distintas de consultar enlaces en Octopus, dependiendo de su tipo: enlace a fuente, enlace a documento y enlaces recursivos. En la búsqueda por enlace, el usuario especifica las características del objeto deseado según sus preferencias, de manera como lo haría en una búsqueda de fuente de datos o de documento, según corresponda.

B. Perfiles de Usuario

Octopus tiene tres tipos de usuario: — Administrador, usuario con más privilegios y responsabilidades. Especifica los predicados por defecto para las búsquedas. Mantiene el catálogo del sistema. — Usuario de Sesión, persona con cuenta en Octopus que ingresa su clave para iniciar sesión. Tiene el privilegio de especificar sus propios términos difusos. Al hacer una búsqueda avanzada, decide usar definiciones por defecto o propias. Puede almacenar resultados de búsquedas. — Usuario General, cualquier internauta que usa Octopus sin iniciar sesión. No tiene privilegio de crear o modificar predicados ni almacenar búsquedas. Los predicados empleados para las búsquedas de este tipo de usuarios son los especificados por el administrador. Puede seleccionar el enlace de registro en la pantalla de inicio y crear una cuenta en Octopus, al finalizar habrá iniciado sesión con sus privilegios.

C. Diseño de la Interfaz

Octopus tiene una interfaz muy similar a la de la mayoría de los buscadores en Internet: campos de texto, botones, botones de opciones, casillas de selección, menús emergentes, así como otros elementos de interfaces ampliamente usados.

La pantalla de inicio de Octopus contiene un sencillo diálogo (Figura 1). En el centro hay una caja de texto en que se colocan las palabras claves de una consulta básica, junto con un botón “Buscar”. Tres palabras “Documentos”, “Fuentes” y “Enlaces” debajo de la caja de texto son una selección del tipo de objeto a buscar. Al lado superior derecho la frase “Log In” permite el ingreso de usuario registrado o la creación de nueva cuenta. El enlace “Ayuda” es de contenido obvio. La frase “Búsqueda Avanzada” es un enlace que lleva a la página de tal tipo de búsquedas. Se usan interfaces de búsquedas visuales que se basa en el paradigma de consultas de bases de datos conocidos

como QBE (Query By Example). Se explica en la sección que describe la funcionalidad de Octopus.



Figura 1: Diálogo Inicial de Octopus

Un elemento importante en la interfaz es el que nos permite la especificación de un predicado en lógica difusa como un rango impreciso. Este tipo de conjuntos difusos es caracterizado con una función de membresía cuya representación gráfica es de forma trapezoidal (Figura 2).

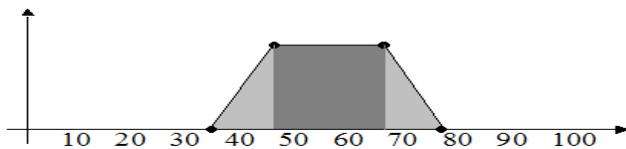


Figura 2: Conjunto Difuso Definido por Trapecio

Por ejemplo, para el predicado “mediano” en un dominio escalar entre 0 y 100, el elemento 55 sin duda tendrá un valor de pertenencia igual 1, en cambio que 0 y 100 tendrán como grado de satisfacción (o membresía) 0; en cambio, podría discutirse el valor de pertenencia de un elemento como 45 ó 37 ó 75, los cuales podrían tener valor de pertenencia diferente de 0 y 1. Habitualmente estos trapecios se especifican por una cuádrupla de valores $(x1, x2, x3, x4)$ donde $x2$ y $x3$ representan el límite inferior y superior del intervalo de total satisfacción (llamado núcleo), mientras $x1$ y $x4$ corresponden al intervalo de no exclusión (llamado soporte). El sistema Octopus se vale de una barra con deslizadores como herramienta de interfaz gráfica para especificar dicho trapecio, como muestra la Figura 3, usando una variación de tonos del color de relleno, además de mostrar los valores numéricos de los límites.

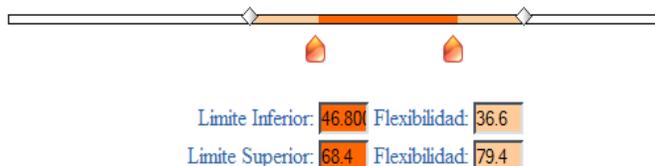


Figura 3: Interfaz Gráfica para Especificar un Trapecio

El gráfico de barra como el de la Figura 3 es un método interactivo para especificar el trapecio. Los rombos blancos indican la flexibilidad que, gráficamente, son los extremos derecho e izquierdo del trapecio que van ubicados sobre el eje de las X. Los íconos anaranjados son empleados para especificar los límites superior e inferior del trapecio que, gráficamente, son los que hacen “el techo” del trapecio y todo lo que haya debajo de ellos y entre los puntos indicados por los rombos (flexibilidad) representa el trapecio en sí. En esta herramienta el usuario puede valerse de los rombos y los

íconos anaranjados para especificar el trapecio, así como modificar los valores de los campos de texto directamente. Como se puede apreciar, el gráfico de barra especificado anteriormente representa el trapecio que se muestra en la Figura 2. El área gris oscura indica una pertenencia del 100%, mientras que los valores que caen en el área gris más clara tienen una pertenencia menor. Como resultado, todas las entradas de una tabla cuyos atributos caigan dentro del área gris oscura serán mostradas, los atributos que caigan dentro de las áreas grises más claras serán mostradas según haya especificado el usuario, y los valores que no estén en el trapecio serán ignorados.

III. FUNCIONALIDAD DE OCTOPUS

Cualquier usuario puede realizar una búsqueda sencilla, simplemente seleccionar el tipo de objeto a buscar (fuente de datos, documento o enlace), luego introduce en un campo de texto las palabras claves y pulsa el botón “Buscar”. Octopus realiza la búsqueda y despliega los resultados por pantalla. La búsqueda avanzada está también disponible para cualquier usuario. Los usuarios registrados que hayan iniciado sesión pueden almacenar los resultados de la búsqueda en el sistema para así poder consultarlos posteriormente, o emplearlos en la búsqueda avanzada para refinar los resultados encontrados. También pueden definir y usar sus propios términos difusos en búsqueda avanzada. A continuación se especifican interfaces y semántica de este tipo de búsqueda, así como otras operaciones que provee Octopus.

A. Especificación de Tipo de Objeto a Buscar

En la búsqueda avanzada es posible involucrar varios objetos de los tipos: fuente de datos, documento y enlace. Para comenzar, el usuario especifica las palabras claves en un campo de texto. Se dispone de tres botones, uno por cada tipo de objeto, los cuales que perimen agregar objetos a la especificación de la búsqueda, como se muestra en la Figura 4. Por cada objeto se seleccionan los atributos que se deseen involucrar en el criterio de búsqueda. Los filtros para cada atributo pueden ser precisos o involucrar términos difusos.



Figura 4: Diálogo Principal de la Búsqueda Avanzada

En la Figura 5 se observan los atributos de un objeto de tipo documento así como la forma de especificar condiciones, usando términos precisos o difusos, según la preferencia del usuario. Por ejemplo, aquí el usuario, entre otras cosas, solicitó que el documento fuese muy nuevo, que el idioma fuese distinto de español y que el tipo de archivo fuese .doc. Claro está, un usuario puede no querer especificar condiciones sobre todos los atributos de la tabla, sino sobre algunos, por eso hay selectores del lado izquierdo de cada atributo para indicar si se considera o no en el criterio de búsqueda.

En los atributos Tamaño y Número_de_Páginas documentos, es posible usar cuantificadores. La interfaz para ello es la mostrada en la Figura 6, la cual va del lado derecho, alineada

con los atributos correspondientes, en el mismo diálogo de la Figura 5, allí no se ven, por razones de espacio.

Documento(d0)		
P.C	Atributo	Opciones de Comparacion:
<input checked="" type="checkbox"/>	Titulo	igual a <input type="text" value="sqlf"/>
<input checked="" type="checkbox"/>	URL	diferente de <input type="text" value="www.google.com"/>
<input checked="" type="checkbox"/>	Autor	igual a <input type="text" value="Tineo"/>
<input checked="" type="checkbox"/>	Idioma	diferente de <input type="text" value="Español"/>
<input checked="" type="checkbox"/>	Tipo de Archivo	igual a <input type="text" value="doc"/>
<input checked="" type="checkbox"/>	Fecha de Creacion	<input checked="" type="radio"/> muy <input type="radio"/> nuevo
<input checked="" type="checkbox"/>	Tamano	grande <input type="radio"/> mediano <input type="radio"/>
<input checked="" type="checkbox"/>	Numero_de_Paginas	escasas <input type="radio"/> muchas <input type="radio"/>

Figura 5: Extracto de Diálogo para Búsqueda Avanzada en Octopus, Indicando Atributos y Condiciones sobre Objetos de Tipo Documento

Opciones de Cuantificadores:		
Mayor	ALL	ejemplo
Menor	ALL	ejemplo

Figura 6: Extracto de Diálogo para Búsqueda Avanzada

En la Figura 6 para la opción de cuantificación “Mayor ALL ejemplo”, la palabra “ejemplo” corresponde al nombre con que se guardaron los resultados de una búsqueda previa. Si el usuario está registrado, podrá colocar aquí alguna de sus búsquedas anteriores que tenga guardadas, y si el usuario no está registrado, podrá indicar aquí resultados de búsquedas anteriores realizadas en su sesión de búsqueda actual, es decir, desde que ingresó al sistema y antes de cerrar su explorador de Internet. Si por el contrario no hay ninguna búsqueda anterior que pueda usarse, esta opción no estará habilitada. La palabra “Mayor” es un comparador, mientras la palabra “ALL” es un cuantificador. En este caso, se trata del cuantificador universal (clásico predefinido), pero es posible seleccionar un cuantificador difuso que sea predefinido del sistema o definido por el usuario, en caso de un usuarios registrado que ha iniciado sesión.

Fuente (f0)		
P.C	Atributo	Opciones de Comparacion:
<input checked="" type="checkbox"/>	Nombre	igual a <input type="text" value="sqlf"/>
<input checked="" type="checkbox"/>	URL	diferente de <input type="text" value="www.google.com"/>
<input checked="" type="checkbox"/>	Pais	igual a <input type="text" value="Venezuela"/>
<input checked="" type="checkbox"/>	Ciudad	igual a <input type="text" value="Caracas"/>
<input checked="" type="checkbox"/>	Unidad de Granularidad	Giga byte
<input checked="" type="checkbox"/>	Recentness	poco <input type="radio"/> antiguo <input type="radio"/>
<input checked="" type="checkbox"/>	Completitud	muy <input type="radio"/> normal <input type="radio"/>
<input checked="" type="checkbox"/>	Frecuencia	muy <input type="radio"/> mensual <input type="radio"/>
<input checked="" type="checkbox"/>	Granularidad	muy <input type="radio"/> media <input type="radio"/>
<input checked="" type="checkbox"/>	OTRO	muy <input type="radio"/> media <input type="radio"/>

Figura 7: Extracto de Diálogo para Búsqueda Avanzada Indicando Condiciones sobre Fuente de Dato

Para objetos de tipo fuente de datos (Figura 7), se proveen interfaces similares a las que se han explicado para los documentos. Para los atributos País, Ciudad, Recentness, Completitud, Frecuencia y Granularidad, se puede usar la opción de cuantificadores, de la manera que se explicó antes. Son los mismos tipos de interfaces y criterios que pueden especificarse, por lo que no vale la pena explicarlos acá, pues sería innecesariamente extenso y redundante.

Existen dos formas de consultar enlaces en el sistema. Si el usuario pulsa el botón “Enlace” y luego el botón “Fuente”, estará solicitando enlaces que tengan entre uno y dos niveles de recursividad hacia algún tipo de fuente. El panel que se le desplegará en pantalla será el mismo que si hubiese pulsado el botón “Fuente” en un principio, y los enlaces mostrados en el resultado serán los que direccionen con uno o dos niveles de recursividad a fuentes que cumplan con las condiciones señaladas. Lo mismo para “Enlace” seguido de “Documento” pero con respecto a Documentos propiamente.

La otra forma de consultar enlaces consiste en primero pulsar alguno de los botones “Fuente” o “Documento” y seguidamente pulsar el botón “Enlace”. Si el usuario primero pulsa “Documento”, luego “Fuente” y seguidamente “Enlace”, se desplegará otro panel de Fuentes en el que el usuario especificará las condiciones que deban cumplir las fuentes que sean direccionadas por ese enlace. Luego puede continuar pulsando cualquiera de los tres botones y agregar nuevos paneles para especificar nuevas instancias de las distintas tablas. Cada vez que se pulse el botón “Enlace”, el panel desplegado dependerá del anterior (documentos o fuentes).

B. Selección y Ordenamiento de Condiciones

En la sección anterior se explicó la primera fase de la búsqueda avanzada en la que el usuario especifica los objetos y condiciones sobre sus atributos. En la siguiente fase, como se muestra en Figura 8, el usuario especifica los atributos a mostrar en la tabla resultante. Así mismo, se indica si se quiere agrupar por algunos atributos.

Tabla Resultado		
<input checked="" type="radio"/> Agrupar	<input type="radio"/> Mostrar	Atributo
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Titulo
<input type="checkbox"/>	<input type="checkbox"/>	URL Documento
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Autor
<input type="checkbox"/>	<input type="checkbox"/>	Idioma
<input type="checkbox"/>	<input type="checkbox"/>	Tipo de Archivo
<input type="checkbox"/>	<input type="checkbox"/>	Fecha de Creacion
<input type="checkbox"/>	<input type="checkbox"/>	Tamano
<input type="checkbox"/>	<input type="checkbox"/>	Numero de Paginas

Figura 8: Extracto de Diálogo para Búsqueda Avanzada en que se Indican Atributos a Mostrar y Atributos de Agregación

Las condiciones que en la fase anterior se especificaron sobre los atributos de los distintos objetos, ahora son combinadas. La combinación involucra los operadores lógicos de negación, conjunción y disyunción, los cuales son interpretados de acuerdo con su semántica en lógica difusa. Finalmente, se puede especificar una condición cuantificada de grupo. La interfaz para ello se muestra en la Figura 9.

Orden de condiciones		
Condiciones		
-	d0.Titulo = "sqlf"	Y
NOT	d0.Idioma = "Español"	Y
-	d0.t_Archivo = "pdf"	-

Y

Figura 9: Extracto de Diálogo para Búsqueda Avanzada en que se Combinan las Condiciones Especificadas

C. Filtrado, Almacenamiento y Exportación de Resultados

En la última fase de la búsqueda avanzada, es posible aplicar varias operaciones sobre los resultados de la misma. Para ello se provee la interfaz mostrada en la Figura 10.

Visualizar la búsqueda junto con otro resultado
Tabla a usar: Fuente Unir Resultados

Visualizar solo elementos que estén en otra tabla resultado
Tabla a usar: Tabla Resultado 1 Intersectar Resultados

Opciones sobre la búsqueda: (Nueva Tabla)
 Sumar: Promedio: Granularidad M.
 Visualizar Opciones

Guardar su Búsqueda
Titulo: Guardar

Desea descargar estos resultados en formato XML
[Click aquí](#)

Figura 10: Extracto de Diálogo para Búsqueda Avanzada en que se Indican Operaciones Finales para los Resultados

Tras varios resultados de búsquedas, el usuario puede aplicar operaciones de conjuntos para refinar aún más la búsqueda. Por ejemplo, en la Figura 10, si el usuario pulsa “Interceptar Resultados” en el ejemplo previo, el sistema mostrará en pantalla el resultado de interceptar el conjunto de resultados de “Tabla Resultado 1” (que es una tabla previamente guardada por un usuario registrado) con el conjunto de los resultados obtenidos tras especificar la última búsqueda avanzada.

En la interfaz de la Figura 10, el usuario también tiene la opción de realizar ciertas operaciones sobre los resultados obtenidos. Por ejemplo, si se pulsa “Visualizar Opciones”, se mostrará el promedio de la granularidad de la Fuente de Datos (de haber consultado una fuente de datos).

También puede observarse en la Figura 10 que el usuario tiene la opción de asignar un título a la búsqueda realizada y guardarla para luego usarla en el futuro. El usuario puede solicitar que se exporten los resultados en un archivo XML (opción de impresión) y descargar dicho archivo para usarlo posteriormente (Figura 10).

D. Operaciones sobre el Catálogo

El administrador se encarga de actualizar, insertar y eliminar elementos en el catálogo de Octopus:

Agregar Fuente de Datos: Al ingresar al catálogo, el administrador verá en pantalla las fuentes de datos presentes en el sistema. El botón “Agregar Nueva Fuente” lleva al administrador a la ventana que le solicita los datos de la nueva fuente a agregar.

Agregar Documento: Insertar un nuevo documento en el sistema y asociarlo a dicha fuente de datos, se le piden los campos respectivos.

Agregar Enlace: Se le solicitará al usuario administrador del sistema que indique una fuente de datos a la cual hacer referencia desde la actual. También se pueden añadir enlaces entre documentos.

Modificar: De vez en cuando es necesario actualizar o corregir los datos de una fuente o documento registrado. Al pulsar este botón el administrador será llevado a la ventana de inserción correspondiente, sólo que esta vez los distintos campos de texto estarán ocupados por los datos insertados anteriormente y queda a juicio del administrador modificarlos según necesite.

Eliminar: Si el administrador lo desea, se puede eliminar una fuente o documento registrado en el sistema. Al pulsar este botón, se mostrará una ventana de alerta consultando al administrador si está seguro o no de querer llevar a cabo esta acción.

Formato XML para agregar una fuente de datos, documentos y enlaces. Los distintos elementos de información del catálogo pueden estar especificados en un documento XML el cual se usaría para agregarlos. El formato general está especificado en un documento XML schema el cual no presentamos aquí por razones de espacio.

IV. TÉRMINOS DIFUSOS DE SQLf EN OCTOPUS

El corazón de las consultas difusas son los términos difusos: predicados, modificadores, comparadores y cuantificadores. Son elementos lingüísticos de naturaleza vaga a los cuales se les da una interpretación basada en la teoría de conjuntos difusos.

Octopus provee de interfaces gráficas para especificar términos difusos, propios de los usuarios de sesión o genéricos dados por el administrador. Adicionalmente permite al administrador que especifique los términos difusos en documentos XML, de acuerdo con lo propuesto por [11]. Los usuarios no tienen que usar directamente la sintaxis de SQLf, aunque Octopus sí la utiliza internamente, pues se implementó con SQLf.

A. Predicados

Los predicados difusos puede decirse que son las condiciones que deben cumplir los distintos atributos de las tablas a fin de flexibilizar las consultas. Por ejemplo, un predicado difuso puede ser “joven”, el cual es un rango de valores en el que entra el atributo “edad” de alguna tabla; entonces, una consulta puede ser buscar los empleados jóvenes de una empresa lo cual, tras definir el predicado “joven”, se traduce en buscar las entradas de la tabla “empleados” tales que el atributo “edad” cumpla con el predicado. Hay tres tipos de predicados difusos en Octopus (Figura 11).

Por Expresión: este tipo de predicados requieren que se especifique un rango de valores (señalando el valor inicial y el final) y una fórmula aritmética para determinar la pertenencia o

“cumplimiento” de la condición del predicado. En SQLf se especifican con una sentencia de la forma: CREATE FUZZY PREDICATE <nombre> ON <dominio> AS <expresión aritmética>.



Figura 11: Extracto de Diálogo para Definir un Predicado

Por Extensión: este tipo de predicados requiere que se especifique cada uno de los valores de un determinado conjunto, así como su grado de pertenencia al conjunto difuso. En SQLf se especifican con una sentencia de la forma: CREATE FUZZY PREDICATE <nombre> ON <domino> AS { <grado1>/<elemento1>, <grado2>/<elemento2>, ... , <gradoN>/<elementoN> }.

Por Trapecio: este predicado requiere que se especifique el trapecio que señala los valores que entran en el área de aceptación, cuáles no y en qué medida. En SQLf se especifican con una sentencia de la forma: CREATE FUZZY PREDICATE <nombre> ON <dominio> AS (x1,x2,x3,x4).

B. Modificadores

Los modificadores son términos difusos que se emplean para crear un nuevo predicado difuso a partir de otro ya existente. La ventaja de los modificadores es que permite llevar a cabo esta operación a la hora de realizar determinada consulta. Matemáticamente, un modificador se encarga de “modificar” el trapecio que determina la pertenencia de un valor a determinado conjunto difuso, pero a la hora de realizar la consulta. Ejemplos de modificadores son “muy”, “pocos”, “demasiados”. La interfaz en Octopus se muestra en la Figura 12. Existen dos tipos de modificadores:

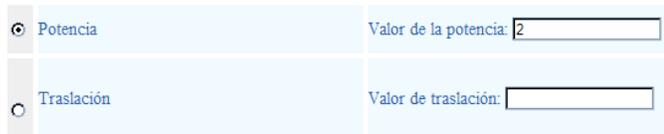


Figura 12: Extracto de Diálogo para Definir un Modificador

Por Potencia: se especifica un valor que “deformará” el trapecio a fin de redefinir el área de pertenencia. En SQLf se especifican con una sentencia de la forma: CREATE FUZZY MODIFIER <nombre> AS POWER <exponente>.

Por Traslación: se especifica un valor que moverá la gráfica de pertenencia a la derecha o a la izquierda, según el valor sea positivo o negativo. En SQLf se especifican con una sentencia de la forma: CREATE FUZZY MODIFIER <nombre> AS TRANSLATION <desplazamiento>.

C. Comparadores

Los comparadores son predicados difusos que permiten comparar de manera flexible, un atributo de una a tabla con un valor que especifique el usuario. Ejemplos de comparadores son “mucho mayor que”, “mucho menor”, “aproximado”, etc. En Octopus existen dos tipos de comparadores (Figura 13):

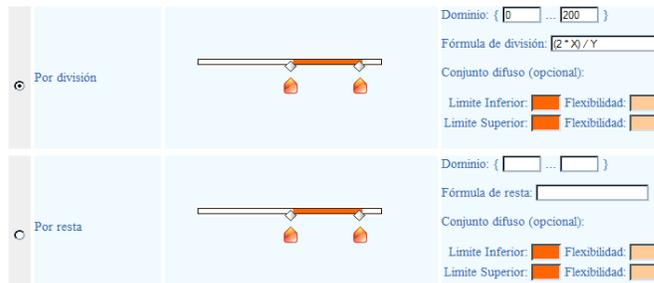


Figura 13: Extracto de Diálogo para Definir un Comparador

Por Resta: se especifica el dominio de valores y una fórmula de resta de elementos para especificar la pertenencia al conjunto difuso. Se puede especificar el conjunto difuso mediante los vértices del trapecio pero es opcional.

Por División: se especifica el dominio de valores y una fórmula de división de elementos para especificar la pertenencia al conjunto difuso. Se puede especificar el conjunto difuso mediante los vértices del trapecio pero es opcional.

D. Cuantificadores

Los Cuantificadores son utilizados para indicar, de cierto conjunto difuso determinado por un predicado, cuál es el subconjunto de valores que se desea obtener.

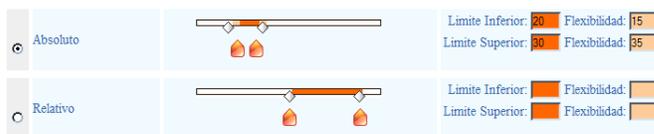


Figura 14: Extracto de Diálogo para Definir un Cuantificador

Ejemplos de cuantificadores predefinidos en el lenguaje SQLf son “cerca de la mitad”, “casi ninguno”, “alrededor de 20”. “casi todos”, “pocos”. En Octopus existen dos tipos de cuantificadores. La interfaz para su definición se muestra en la Figura 14.

Relativos: caracterizan proporciones del número de elementos que satisfacen una condición con respecto al total de elementos en el universo o dominio de la condición. Se definen a través de conjuntos difusos en el intervalo real [0,1]. En SQLf se especifican con una sentencia de la forma: CREATE PROPORTIONAL FUZZY QUANTIFIER <nombre> AS (x1,x2,x3,x4).

Absolutos: caracterizan el número de elementos que satisfacen cierta condición, independientemente de cuál sea la cardinalidad del conjunto base. Se definen a través de conjuntos difusos sobre los números no negativos. En SQLf se especifican con una sentencia de la forma: CREATE ABSOLUTE FUZZY QUANTIFIER <nombre> AS (x1,x2,x3,x4).

V. CONCLUSIONES Y TRABAJOS FUTUROS

Se ha presentado Octopus, una herramienta que soluciona el problema de selección de fuente de datos y documentos Web utilizando el poder de la lógica difusa. Esta herramienta se ha desarrollado el uso de SQLf a través de un sistema de consulta difuso construido sobre un RDBMS concebido para el desarrollo de aplicaciones de Internet. Las capacidades de consultas difusas de son útiles para expresar las preferencias de los usuarios sobre los criterios de calidad y selección de fuentes de datos con la discriminación entre ellos. Expresión de esas preferencias sería muy difícil consultar con los lenguajes tradicionales de bases de datos. Octopus trabaja con un catálogo para el registro de documentos y fuentes de datos con sus parámetros o criterios de calidad. Administra registros de los documentos publicados y las fuentes de datos. El usuario define los parámetros de calidad y términos difusos que se utilizarían para la selección. Una característica de Octopus es el asistente para la consulta difusa para búsquedas avanzadas usando el paradigma QBE (Query By Example). Este asistente consta de una simple interfaz de usuario final y un mecanismo que construye las consultas difusas en SQLf. Octopus permite la especificación del catálogo de fuentes de datos y los términos difusos a través de XML, asimismo, puede producir sus resultados en XML, de manera que está capacitado para funcionar como un servicio Web.

Una funcionalidad interesante que podría ser añadido a esta herramienta, en algún trabajo futuro, es la conexión con una herramienta de descubrimiento de conocimientos o la aplicación de la minería de texto. Podríamos proponer un motor que realiza el registro automático y la anotación de los documentos y las fuentes de datos. Podría ser objeto de un nuevo desarrollo para complementar Octopus. Problema de la rigidez tradicional (CRISP) herramientas de búsqueda se caracteriza por la respuesta vacía o más pobladas conjunto de respuestas. Lógica difusa resuelve este problema, pero, en algunos casos el número inadecuado de la respuesta parece. No son técnicas destinadas a hacer frente a este problema. La incorporación de estas técnicas sería una importante contribución al proyecto de desarrollo de Octopus. Aquí no hemos tratado con el problema de rendimiento de consulta difusa y sistemas de recuperación de información. Sin embargo, la consulta sistema utiliza una estrategia de evaluación sobre la base de la distribución de un corte. Este mecanismo se conoce como el Principio de Derivación y ha demostrado que poseen el mejor rendimiento entre los mecanismos de evaluación propuestos. Un estudio formal de ejecución sería cuestión de trabajo futuro. Sería muy

interesante hacer una comparación completa de Octopus y el rendimiento con respecto a herramientas similares basadas en búsquedas con criterios precisos clásicos.

AGRADECIMIENTOS

Hacemos un reconocimiento a nuestros apreciados alumnos, y ahora colegas, que colaboraron en el Taller de Desarrollo de Software de la Universidad Simón Bolívar para la construcción de Octopus: Gracias muchachos por disponerse a dar su aporte voluntario a nuestro trabajo, que el Señor les recompense. “*Toda buena dádiva y todo don perfecto descende de lo alto, del Padre de las luces, en el cual no hay mudanza, ni sombra de variación.*” (Santiago 1:17)

REFERENCIAS

- [1] P. Bosc and O. Pivert, *SQLf: A Relational Database Language for Fuzzy Querying*, IEEE Transaction on Fuzzy Systems, vol. 3, no. 1, pp. 1-17, 1995.
- [2] J. Galindo, *New Characteristics in FSQl, a Fuzzy SQL for Fuzzy Databases*, WSEAS Transactions on Information Science and Applications 2, vol. 2, pp. 161-169, 2005.
- [3] J. Galindo, *Handbook of Research on Fuzzy Information Processing in Databases*. 1st edition, Hershey, PA, USA: Information Science Reference, 2008.
- [4] C. González, M. Goncalves, and L. Tineo, *A New Upgrade to SQLf: Towards an Standard in Fuzzy Databases*, in proceedings of the 20th International Workshop on Database and Expert Systems Applications, pp. 442-446, Linz, Austria, September 2009.
- [5] O. Pivert and P. Bosc, *Fuzzy Preference Queries to Relational Databases*, 1st edition, Imperial College Press, 2012.
- [6] M. Goncalves and L. Tineo, *WWW Data Source Selection with SQLf*, in proceedings of the 15th Fuzz-IEEE Conference, Reno, Nevada, USA, May 2005.
- [7] J. Eduardo, M. Goncalves, and L. Tineo, *A Fuzzy Querying System Based on SQLf2 and SQLf3*, in proceeding of the XXX Latin American Conference on Informatics (CLEI 2004), Arequipa, Perú, September 2004.
- [8] M. Goncalves and L. Tineo, *SQLfi and its Applications*, Revista Avances en Sistemas e Informática, vol. 5, no. 2, pp. 33-40, 2008.
- [9] A. Mendelzon, G. Mihaila, and T. Milo, *Querying the World Wide Web*, Journal of Digital Libraries, pp. 68-88, 1997.
- [10] G. Mihaila, L. Raschid, and A. Tomic, *Equal Time for Data on the Internet with WebSemantics*, in proceedings of the 6th International Conference on Extending Database Technology (EDBT'98), Valencia, Spain, March 1998.
- [11] B. Oliboni and G. Pozzani, *Representing Fuzzy Information by Using XML Schema*, in proceedings of the 19th International Workshop on Database and Expert Systems Applications (DEXA 2008) and Third International Workshop on Flexible Database and Information System Technology. IEEE Computer Society Press, pp. 683-687, Turin, Italy, September 2008.

Identificación de Secuencias de Scanning en Redes 802.11

Antonio Araujo¹, Andrés Arcia-Moret²
aaraujo@cenditel.gob.ve, aarcia_m@ictp.it

¹ Centro Nacional de Desarrollo e Investigación en Tecnologías Libres (CENDITEL), Mérida, Venezuela

² Marconi Laboratory, International Centre for Theoretical Physics (ICTP), Trieste, Italia

Resumen: Actualmente las redes 802.11 se han convertido en la primera opción para el acceso inalámbrico a la Internet. Para conectarse, un dispositivo cualquiera se registra a algún Punto de Acceso pasando primero por un costoso proceso de búsqueda de los APs disponibles. El estándar IEEE 802.11 deja abierta la posibilidad de implementar los criterios de búsqueda estableciendo solamente los mensajes necesarios para tal efecto. En este artículo caracterizamos distintos dispositivos a través de un sistema especialmente construido para tal fin. El sistema propuesto es capaz de escuchar todos los canales de manera simultánea para así mostrar verazmente las estrategias que siguen distintos dispositivos durante el proceso de búsqueda. El sistema propuesto permite movilidad, operación bajo estándares abiertos y uso de componentes de bajo costo.

Palabras Clave: IEEE 802.11; Scanning Activo; Caracterización.

Abstract: Nowadays 802.11 networks are the first option to last-mile wireless access to Internet. Access Points operate in a set of overlapping channels conforming a rather chaotic topology. And so, connecting to an available AP requires an expensive scanning process before deciding which AP to connect to. As the IEEE 802.11 do not address specifics on scanning, vendors implement different (closed) strategies for scanning. In this paper we discuss the building of a simple multi-channel sniffer using open standards and low cost components in order to identify scanning sequences. Then, we characterise several scanning strategies within different chipsets.

Keywords: IEEE 802.11; Active Scanning; Characterization.

I. INTRODUCCIÓN

Las redes inalámbricas constituyen una de las principales tecnologías de acceso local y móvil a la Internet. El uso de clientes inalámbricos, que van desde sensores hasta computadores personales, requieren que las redes 802.11 den soporte a las necesidades de usuarios que demandan movilidad y mejor calidad de servicio. Estas demandas pueden verse satisfechas a través de los mecanismos de asociación de los dispositivos móviles a las redes 802.11, cuando estos pueden orientarse a minimizar los tiempos de búsqueda de Puntos de Acceso disponibles.

El proceso de búsqueda o scanning, ha sido estudiado desde diversos enfoques para disminuir los tiempos de conexión de un dispositivo móvil. Desde la modificación de variables establecidas en el estándar IEEE 802.11 [1], hasta la incorporación de información adicional a las tramas del estándar [2], las propuestas consideran distintas maneras de mejorar el algoritmo de scanning de los dispositivos. La identificación de secuencias de scanning permite comprender el comportamiento de distintos dispositivos WiFi de uso masivo, los que hasta hoy, no muestran el código fuente de la implementación del scanning. La mejora del proceso de scanning podría, a su

vez, representar directamente una mejora en el proceso de *handover* (el cual ocupa hasta un 80% del tiempo de reconexión [3]), pues los clientes en movimiento buscan constantemente reconectarse.

Para entender el tráfico real de una red, se utiliza un analizador de paquetes o sniffer que captura tramas a través de la interfaz de red. En el caso de las redes 802.11, un sniffer común captura el tráfico de red en un canal particular de operación; esto es, en uno de los 11 canales de la banda de 2.4GHz. Debido a que los dispositivos inalámbricos ejecutan el proceso de scanning enviando tramas para detectar puntos de acceso en todos los canales, un sniffer con una sola interfaz de red no es suficiente para capturar todo el tráfico generado durante el proceso de scanning. Algunas aplicaciones de software como Kismet [4] permiten capturar tramas en redes inalámbricas 802.11 en los 11 canales. El proceso de captura se realiza en intervalos de tiempo ajustables en los que el sniffer permanece en un canal particular y luego salta a otro hasta recorrer todos los canales. A pesar de hacer un barrido por todos los canales de la banda 2.4GHz es posible que algunas tramas no sean capturadas mientras se realiza el salto entre canales. Regularmente, con éste enfoque, se está capturando menos de

un 10% del tráfico presente en los 11 canales, pues se puede examinar 1 canal a la vez.

En este trabajo se presenta el diseño y prueba de un sniffer de redes 802.11 que pueda escuchar, simultáneamente, los 11 canales de la banda 2.4GHz. Se prueba experimentalmente que es viable realizar un proceso de ingeniería inversa de los algoritmos de scanning implementados en dispositivos con distintos sistemas operativos para dispositivos móviles. También, se establecen algunas premisas de diseño para la construcción del prototipo y finalmente, se realizan experimentos para una caracterización de los algoritmos de scanning.

A. El Proceso de Scanning en Redes 802.11

El *scanning* de las redes inalámbricas es una función importante del protocolo MAC IEEE 802.11, donde los nodos intentan buscar redes inalámbricas disponibles y luego intentan asociarse a ellas [5]. El propósito del procedimiento de *scanning* es encontrar todos los BSS (Basic Service Set) a los que la estación pueda unirse. Para ser capaz de procesar todos los puntos de acceso en un área específica, por lo general se ajusta un timer llamado MaxChannelTime en áreas con un gran número de puntos de acceso [6].

En el estándar IEEE 802.11 [7] se definen dos tipos de técnicas de scanning de redes: scanning activo y scanning pasivo.

- *Scanning activo*: implica la generación de tramas Probe Request y el posterior procesamiento de las tramas Probe Response recibidas. Un resumen del procedimiento haciendo énfasis en los eventos de interés es el siguiente:
 - Enviar una trama Probe Request con destino broadcast.
 - Iniciar un temporizador (ProbeTimer) a la espera de respuestas.
 - Si el medio no está ocupado antes de que ProbeTimer alcance el valor de MinChannelTime, pasar a escanear el siguiente canal.
 - En caso contrario, activar un nuevo temporizador (MaxChannelTime) y cuando expire, procesar todas las tramas Probe Response para pasar al siguiente canal.
- *Scanning pasivo*: Si el parámetro ScanType indica un scanning pasivo, la estación escucha a cada canal una duración máxima definida por el parámetro MaxChannelTime. Los dispositivos móviles generalmente implementan el algoritmo de scanning activo para el descubrimiento de Puntos de Acceso (AP por sus siglas en inglés).

Particularmente, en este trabajo extraemos nuestro análisis y conclusiones basados en el scanning activo, pues de los dos procesos, es el que puede controlarse para obtener una eventual reducción de los tiempos de conexión a una red 802.11.

El artículo está estructurado de la siguiente manera. La Sección II discute el estado del arte sobre scanning en redes 802.11. La Sección III propone el diseño de un sniffer automático para el scanning en redes 802.11 con su respectiva arquitectura. La Sección IV discute el diseño experimental y el proceso de medición. La Sección V muestra los distintos algoritmos de scanning encontrados en una muestra de dispositivos

móviles. Finalmente, la Sección VI presenta las conclusiones y el trabajo futuro.

II. TRABAJOS RELACIONADOS

El estándar IEEE 802.11 describe el *scanning* para dispositivos inalámbricos (así como en la sección anterior), pero no especifica ni el orden particular ni los tiempos a esperar que deben seguirse para enviar las tramas de administración *Probe Request* en los canales del espectro. En consecuencia, así como lo prueban los distintos experimentos realizados, los fabricantes de tarjetas de red inalámbricas pueden implementar secuencias arbitrarias de *scanning* en sus algoritmos directamente en el hardware, así como también permitir que los sistemas operativos definan una secuencia particular para el *scanning*. La razón de la escogencia particular de una secuencia de scanning es hasta ahora una pregunta abierta.

Gupta, Beyah y Corbett [5] describen un proceso empírico para caracterizar algoritmos de scanning de diferentes tarjetas inalámbricas. Se realiza un proceso de scanning completo para probar todos los canales del espectro 2.4GHz y para utilizar un algoritmo que genera una sola traza de captura de tramas. La caracterización de interfaces inalámbricas se realiza en base a los canales en los cuales se envía la primera trama Probe Request, el número de tramas Probe Request enviadas por cada canal, ráfagas de tramas Probe Request y la cantidad de tiempo de espera de la interfaz inalámbrica en cada canal del espectro 2.4GHz. Los autores proponen la caracterización para ayudar al entendimiento de las diferentes implementaciones de scanning activo, tanto en software como en hardware, así como servir de base para evaluar su comportamiento en simuladores de redes.

Kim y Kim [2] estudian el desempeño del scanning de canales en redes IEEE 802.11e, redes en las que se define un conjunto de mejoras de calidad de servicio para aplicaciones inalámbricas a través de modificaciones de la capa de control de acceso al medio (MAC por sus siglas en inglés). La propuesta se basa en un scanning inteligente que considera diferentes prioridades y analiza el desempeño del scanning de canales con respecto al número de estaciones activas.

Laurenson [8] propone el diseño de un sistema para adquisición y preservación de tráfico de red inalámbrico. El autor revisa varios enfoques para la recolección de tráfico útil en procedimientos de análisis forense digital que garanticen su confiabilidad. El sistema propuesto está compuesto por sensores inalámbricos, representados por Puntos de Acceso con software modificado, para recolectar y enviar tráfico de red a un servidor forense centralizado que almacena y preserva los datos adquiridos en un entorno WLAN existente.

Corbett, Beyah y Copeland [9] emplean el procesamiento de señales para analizar la periodicidad del tráfico generado por el scanning con un análisis espectral. El mecanismo puede ser utilizado para detectar sistemas no autorizados que usan tarjetas de interfaz de red diferentes de las reconocidas como válidas. Se muestra que pueden distinguir tarjetas de diferentes fabricantes a través de un perfil espectral.

Reddy, Sharpe y Paulraj [10] proponen un sistema en el que se pueden escanear los canales en la banda de radio de 2.4GHz y múltiples canales en el espectro inalámbrico de 5GHz simultáneamente. Los autores utilizan una arquitectura basada en dos subsistemas: un computador personal y un conjunto de tarjetas de computador del tipo Single Board Computer, SBC (por sus siglas en inglés). Cada una de ellas con soporte para múltiples tarjetas inalámbricas en formato Mini PCI. Estos subsistemas se comunican a través de interfaces Ethernet. Un software específico se ejecuta en el subsistema Host y controla las tarjetas inalámbricas conectadas en cada SBC para realizar el proceso de escaneo. El uso de un computador Host (PC de escritorio) y un Switch Ethernet para su interconexión, hacen al sistema “estacionario” o de escritorio y de costos elevados.

III. SNIFFER MULTICANAL PORTATIL

El comportamiento de las interfaces inalámbricas 802.11 en los diferentes dispositivos móviles sigue las especificaciones del estándar IEEE 802.11. El proceso de scanning activo está definido por ese estándar, sin embargo los fabricantes de los dispositivos pueden implementar distintas versiones del algoritmo, tanto en software, como hardware, para encontrar puntos de acceso disponibles más o menos rápido (como veremos más adelante en la Sección V). Sin embargo, Castignani, Arcia-Moret y Montavont [1] muestran que no solamente es de interés encontrar rápidamente los APs disponibles, sino que para llevar un proceso de descubierta de calidad, también es necesario tomar en cuenta la cantidad de APs que dependa básicamente del tráfico circulante en el momento del descubrimiento.

En esta sección mostramos el diseño y construcción de un dispositivo automático para realizar ingeniería inversa del proceso de scanning de clientes en redes 802.11. En este estudio el sniffer actúa como elemento adicional e inocuo para el desempeño de una red 802.11 y que permitirá recolectar datos de las estaciones móviles y analizar la manera en que realizan el proceso de scanning activo. La Figura 1 muestra un bosquejo de una red 802.11 con un sniffer automático. Como se puede observar, el sniffer debe ser capaz de recoger los Probe Request y Probe Response de diferentes canales y reportar, correctamente, la estampilla de tiempo de la trama escuchada.

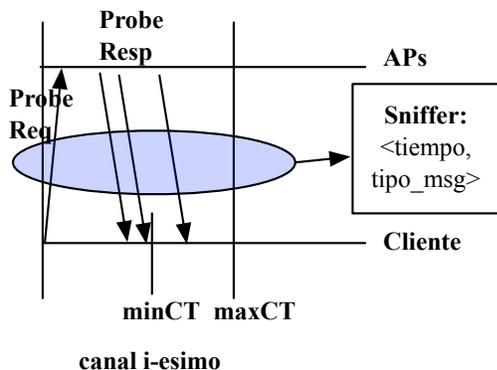


Figure 1: Bosquejo de Red IEEE 802.11 con un Sniffer

En el sistema propuesto para redes 802.11, a diferencia de los sniffer de red comunes, se puede escuchar tráfico en los 11 canales de la banda 2.4GHz simultáneamente. En la Figura 2 se muestra un diagrama de bloques con la arquitectura del sniffer propuesto. Cada uno de los componentes se describe a continuación.

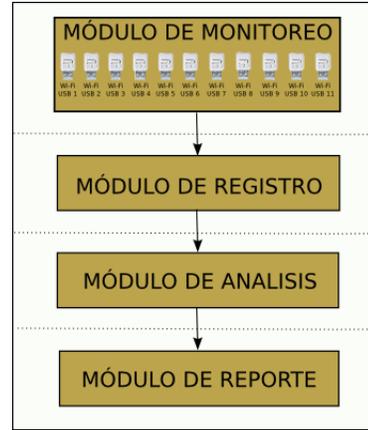


Figure 2: Diagrama de Bloques de la Arquitectura de Sniffer Propuestos

A. Premisas de Diseño

El sniffer se diseñó siguiendo las siguientes premisas:

- Usar componentes de fácil acceso y bajo costo. La facilidad para conseguir los elementos que forman parte del sniffer permitirá su rápida construcción así como la reproducción de cualquier estudio que se proponga de manera sencilla.
- Construir el dispositivo que sea móvil y tenga autonomía de energía. La posibilidad de movilizar el sniffer con autonomía en distintos ambientes permitirá la realización de estudios de dispositivos clientes en redes con distintas configuraciones y concurridas por diferentes tipos de clientes.
- Operar bajo estándares abiertos. Se promueve el uso de estándar y tecnologías abiertas y libres. De esta manera se puede lograr un estudio preciso del funcionamiento de los algoritmos de scanning en los clientes.

B. Módulos del Sistema

A continuación se describe la función específica de cada módulo. Cada uno representa una pieza de software separable según su función, asegurando un mínimo acoplamiento entre ellos.

1) *Módulo de Monitoreo*: El módulo de monitoreo se encarga de habilitar y deshabilitar adecuadamente los dispositivos inalámbricos para la captura de tramas en el sniffer. El dispositivo de captura de tramas que se propone es un adaptador externo, inalámbrico WiFi USB. Como la propuesta del sniffer está orientada al estudio en el espectro de 2.4GHz de las redes 802.11, el módulo debe tener la capacidad de gestionar 11 dispositivos inalámbricos correspondientes a los 11 canales. Cada adaptador debe estar soportado por el entorno operativo

del sniffer para habilitar su funcionamiento correctamente. La captura de tramas en cada canal ocurre luego de habilitar el modo monitor en cada adaptador inalámbrico [11]. Un proceso automático habilitará o deshabilitará los adaptadores de acuerdo al entorno operativo que utilice el sniffer.

2) *Módulo de Registro:* La función del módulo de registro es activar y desactivar la captura de tramas en los adaptadores inalámbricos habilitados por el módulo de monitoreo. Se generan archivos separados para cada uno de los canales en el que opera el adaptador (i.e., cada captura es atendida por un proceso por separado). Cada trama registrada es almacenada como una entrada en el archivo bajo el formato pcap [12]. El tipo de trama se obtiene a partir de la información del encabezado radiotap [13]. Los archivos generados por cada canal son integrados en un solo archivo (compactado a partir de varias fuentes) que luego pasa a ser procesado en el módulo de análisis. Para hacer la captura eficiente (por ejemplo, en medio de alto tráfico), se utilizan filtros que establecen la información asociada al estándar 802.11 que se desea almacenar en los registros. La información que se puede registrar para cada trama está limitada por la herramienta soportada en el entorno operativo del sniffer.

3) *Módulo de Análisis:* La función del módulo de análisis consiste en aplicar algoritmos de procesamiento sobre los archivos de tramas capturadas para los 11 canales del espectro de 2.4GHz. Los algoritmos permiten realizar cálculos, generar datos y estadísticas para responder preguntas sobre el comportamiento del proceso de scanning de estaciones móviles. Los algoritmos de análisis pueden ser implementados por programas o scripts soportados en el entorno operativo del sniffer.

4) *Módulo de Reporte:* La función de éste módulo consiste en generar reportes en forma de textos, gráficos y/o animaciones que puedan representar características del comportamiento de scanning activo. Si bien se trata de un módulo que no es obligatorio para llevar a cabo la función principal del sistema, puede ser usado extemporáneamente sobre los datos obtenidos en el módulo anterior. Los reportes pueden ser implementados en programas scripts soportados en el entorno operativo del sniffer y que podrían apoyarse en algún tipo de interfaz de usuario.

IV. DISEÑO EXPERIMENTAL

Sobre la base del diseño y premisas descritas en la sección III se construye un prototipo experimental cuyos componentes se muestran en la Figura 3.

La captura de tramas en los 11 canales de la banda 2.4GHz se realiza a través de adaptadores inalámbricos WiFi USB. Los adaptadores están conectados a un par de concentradores de puertos USB que a su vez se conectan a dos puertos USB del computador portátil.

Originalmente, el sistema empezó a ser diseñado en base a dispositivos de bajo costo como la Raspberry Pi [14]. Sin embargo, hemos constatado que el límite en capacidad de procesamiento hace poco escalable la adquisición de tramas en distintos canales en paralelo, así como es posible en



Figure 3: Componentes del Prototipo Experimental de Sniffer

Table I: Adaptadores Inalámbricos WiFi USB Utilizados

Marca	Modelo	Cantidad	Estándares
TPLINK	TLW723N	8	IEEE 802.11 b/g/n
TENDA	W311M	2	IEEE 802.11 b/g/n
TENDA	W311MI	1	IEEE 802.11 b/g/n

un computador portátil. La diferencia en velocidad de CPU es de 1:3 (700MHz:2.2GHz) y en memoria RAM de 1:4 (512MB:2GB). Respecto al almacenamiento, la Raspberry Pi utiliza una SD card a diferencia del computador portátil que utiliza un disco duro. Para una SD card clase 4 (class 4) el acceso de lectura y escritura mínimo comparado con un disco duro de 5400 RPM es de aproximadamente 1:37 (4MB/s para la SD card [15] y 150MB/s para el disco duro [16]).

El uso compartido de los puertos USB restringe la tasa de transferencia de datos teórica a 480 Mbit/s por cada controlador USB 2.0 en un computador [17]. En los experimentos realizados no se ha encontrado evidencia de que el uso de concentradores USB genere un cuello de botella artificial que limite el registro de tramas de administración del estándar IEEE 802.11.

A diferencia de la propuesta de Reddy, Sharme y Paulraj [10], el diseño propuesto en este documento utiliza elementos de hardware de fácil acceso, bajo costo y hace énfasis en la portabilidad del equipo para recolectar datos e identificar cómo los diferentes sistemas operativos realizan el proceso de scanning en redes 802.11.

Para la construcción del sniffer multicanal se utilizaron los siguientes equipos:

- Computador portátil: Dell Inspiron 1420 con sistema operativo Backtrack Linux [18] con kernel 3.2.6, 2 GB de memoria RAM y disco duro SATA de 160 GB a 5400 RPM.
- 2 concentradores USB con 7 puertos.
- 11 adaptadores de red WiFi USB. En la Tabla I se muestran las características de los adaptadores.

Sin contar el computador portátil, el costo total para construir este prototipo (fácilmente portable) es de aproximadamente 164 US\$.

Table II: Dispositivos Clientes Utilizados para las Pruebas

Dispositivo	Sistema Operativo
Teléfono Nokia N950	Meego 1.2 Harmattan
Samsung Galaxy Tab	Android 3.2
Laptop Dell Inspiron	Debian GNU/Linux 6.0
Laptop HP	Microsoft Windows 7 Ultimate
Apple iPhone 4S	iOS 6.1.2
AP Linksys WRT54G	-

A. Experimentación

Para analizar y caracterizar el comportamiento de los algoritmos de scanning en redes 802.11 se realizan experimentos con el sistema descrito en la sección IV y dispositivos móviles con diferentes sistemas operativos. Como reportan Castignani, Arcia-Moret y Montavont [1], no existe una secuencia óptima predefinida ni un tiempo fijo para ejecutar la secuencia. En nuestro mejor esfuerzo de búsqueda, podemos confirmar que no está publicada una secuencia óptima de búsqueda. En la Tabla II se listan los dispositivos evaluados como clientes y un AP utilizados para pruebas del sniffer experimental.

1) *Elaboración de la Trazas:* La función de monitoreo del sniffer experimental se realiza a través de la suite de aplicaciones *Aircrack-ng* [19], específicamente con la utilidad de línea de comando *airmon-ng* que habilita el modo monitor en cada interfaz inalámbrica para escuchar todos los paquetes del estándar IEEE 802.11.

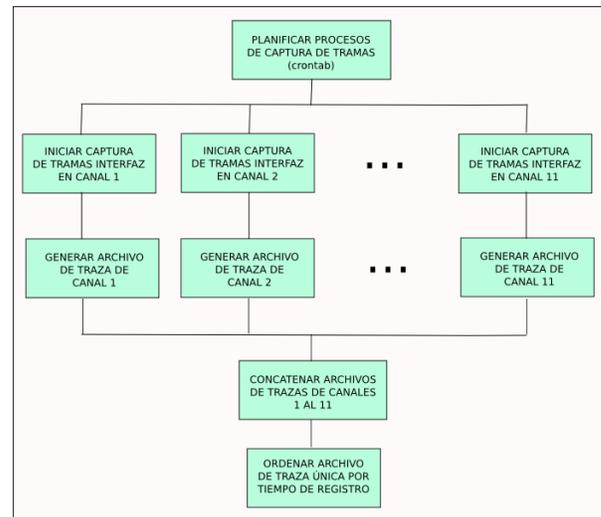
La función de registro del sniffer experimental se realiza con la utilidad de línea de comando *tshark* del analizador de protocolo de red *Wireshark* [20]. *tshark* permite capturar tramas y distintos campos asociados al tráfico de redes alámbricas e inalámbricas. La salida de *tshark* se almacena en un archivo que mantiene la traza asociada a la captura de cada uno de los 11 canales de la banda 2.4GHz.

El prototipo experimental de sniffer planifica la ejecución del comando *tshark* para cada una de las interfaces inalámbricas; ésto a través de la utilidad de línea de comando *crontab* [21]. Se planifica la ejecución del comando *tshark* en el mismo instante de tiempo para cada una de las once (11) interfaces inalámbricas asociadas a los canales de la banda 2.4GHz. Esto permite que la captura de tramas mantengan la misma línea de tiempo de referencia y se capture el comportamiento de clientes en todos los canales por un período determinado.

Un proceso automático (tipo script) realiza la concatenación de cada archivo de captura de canal uno tras otro, desde el canal 1 hasta el canal 11, para obtener un archivo con todas las tramas capturadas. El archivo generado de la concatenación mantiene los mismos campos de un archivo de captura de canal, se emplean utilidades de línea de comando para procesar el archivo y realizar un proceso de ordenamiento de los registros sobre la base del tiempo de captura. La Figura 4 muestra un esquema de construcción del archivo de traza única de captura de tramas.

2) *Duración del Scanning Activo:* A continuación se describen los pasos realizados para la medición del proceso de scanning activo. En el cliente:

- Borrar el caché de las redes inalámbricas de la estación.

**Figure 4:** Esquema de Construcción de Archivo de Trazas Única

Pues, cada vez que el sistema operativo se inicia o despierta luego del proceso de hibernación, realiza un scanning de todos los AP disponibles. Si encuentra un AP que el usuario había seleccionado como “preferido” anteriormente, automáticamente se asocia a él y no realiza el scanning completo en todos los canales [22].

- Iniciar el proceso de scanning en el cliente al activar la conexión inalámbrica.

Del lado del Sniffer:

- Deshabilitar procesos que gestionan servicios de red como networkmanager y bluetooth.
- Configurar los adaptadores inalámbricos para ejecutar el modo monitor en cada uno de los 11 canales. Esto habilita el módulo de Monitoreo del sniffer como se describe en la subsección III-B1.
- Ejecutar el programa *tshark* en cada una de las interfaces en modo Monitor para iniciar la captura de tramas. Este paso habilita el módulo de Registro del sniffer como se describe en la subsección III-B2.
- Esperar la captura (para nuestro caso particular, 4 minutos eran suficientes).
- Detener la ejecución de *tshark* en todas las interfaces y generar el archivo de tramas únicas asociado al scanning en los 11 canales. Estas actividades las realiza el módulo de Registro del sniffer como se describe en la subsección III-B2.

V. RESULTADOS

En esta sección presentamos los dos resultados principales obtenidos. Primero, mostramos las secuencias de scanning de cada uno de los distintos clientes estudiados. Segundo, presentamos la frecuencia con la que se ejecuta un proceso de scanning en cada móvil. Notamos que se trata en cada caso de un proceso bien particular, propio de cada fabricante.

Las tramas capturadas para todos los clientes, tienen la siguiente estructura:

- *epoch_time*: tiempo en segundos transcurridos desde el 1 de enero de 1970 hasta el momento de captura de la trama.
- *MAC origen*: dirección MAC origen del dispositivo que origina la trama.
- *MAC destino*: dirección MAC destino de la trama.
- *Número de secuencia*: número de secuencia (SN) de la trama.
- *Tipo de trama*: subtipo de la trama de acuerdo al estándar IEEE 802.11. En este caso corresponde al valor 4 para tramas ProbeRequest.
- *SSI Signal*: Potencia de la señal recibida (dBm)
- *Canal*: número del canal en el que se registra la trama.

Los archivos consolidados de captura de tramas registran los números de secuencia de cada trama escuchada y su correspondiente tiempo de captura. Con estos datos es posible establecer un orden único a los 11 archivos generados por el sniffer cuando se escucha el tráfico, simultáneamente, en todos los canales.

A. Las Secuencias de Scanning

El estándar 802.11 define el algoritmo de *scanning activo* pero no especifica el orden en el cual la tarjeta inalámbrica de un dispositivo móvil debe probar cada uno de los canales. Se podría pensar que, en la banda de 2.4GHz, los canales del 1 al 11 se prueban de forma secuencial. La posibilidad de capturar tramas de los 11 canales del espectro 2.4GHz permite observar la secuencia de scanning de una STA a partir de los archivos de trama única, como los que se obtienen con el prototipo de sniffer construido.

Una representación de la secuencia de scanning se puede obtener al graficar el canal en el que se recibe la trama en función del tiempo. En esta representación es posible superponer capturas de tramas correspondientes a secuencias completas de scanning o full scanning en las cuales se observan algunos patrones.

En las Figuras 5, 6, 7, 8 y 9 se muestran gráficas superpuestas de los canales en que se capturan tramas con respecto al tiempo de duración de 4 full scanning para cada uno de los dispositivos evaluados. Cada punto de la gráfica representa, el tiempo en el que una trama Probe Request es recibida en un canal particular del espectro 2.4GHz (entre 1 y 11).

En general, en las pruebas realizadas se observó una búsqueda lineal en el sondeo de canales que van desde los números más bajos a los más altos. A medida que se sube de los números más bajos se prueban los canales vecinos próximos generando un recorrido zigzagueante. Estos recorridos hacen una búsqueda exhaustiva de APs en todos los canales. Creemos que este patrón de búsqueda es simple y no considera la escucha de un mismo Probe Request en diferentes canales. Esta situación sugiere que la búsqueda podría optimizarse. Sin embargo, observamos también que, en la Figura 9 el iPhone 4S muestra un comportamiento particular que no se había notado en ninguno de los otros dispositivos evaluados. Para éste móvil en particular, inicialmente, existe una fase

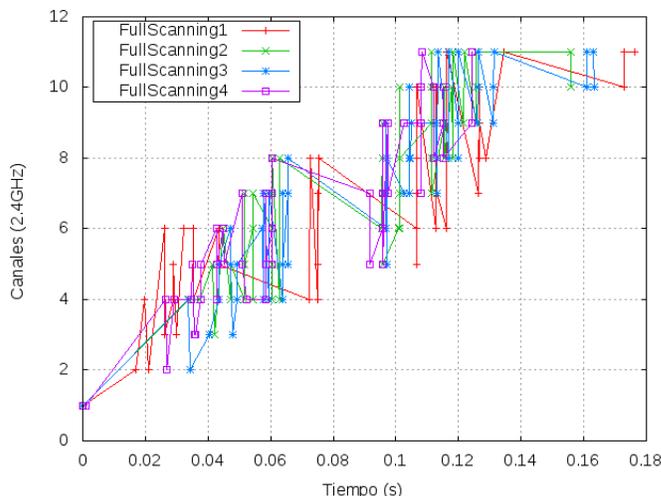


Figure 5: Secuencias de Full Scanning para Nokia N950

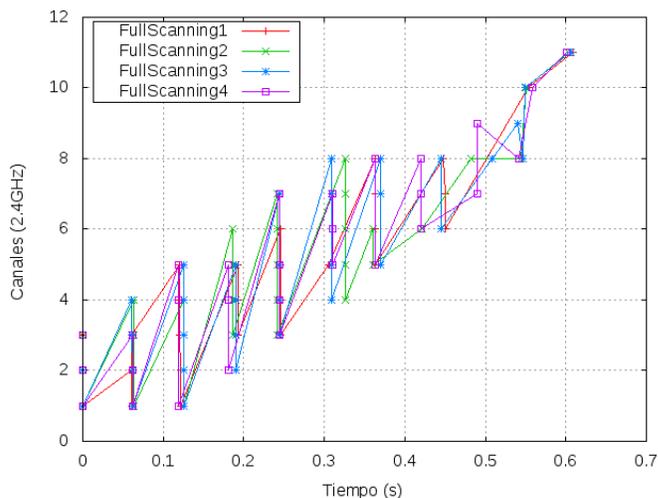


Figure 6: Secuencias de Scanning para Samsung Galaxy Tab

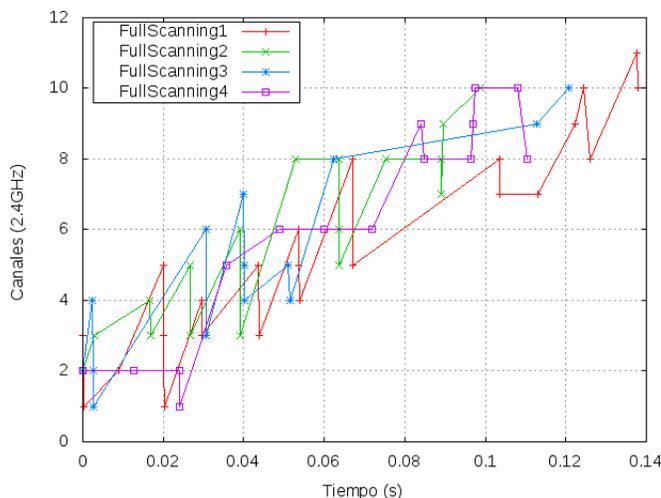


Figure 7: Secuencias de scanning para Laptop Dell

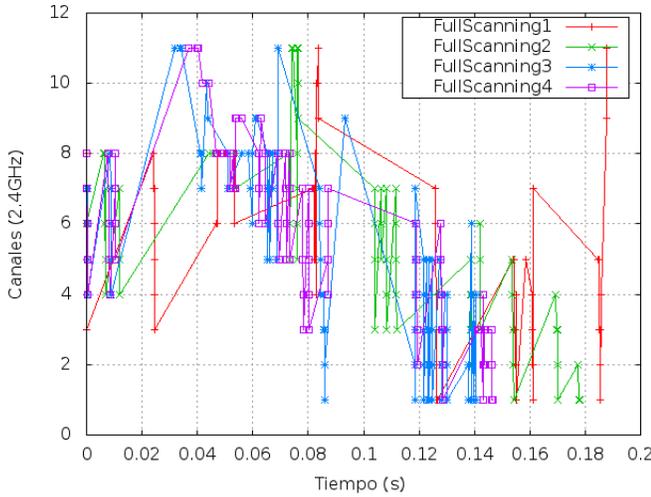


Figure 8: Secuencias de *Scanning* para Laptop HP

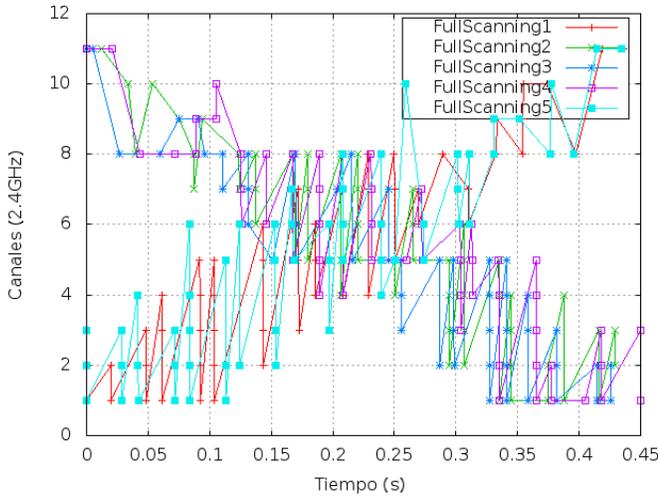


Figure 9: Secuencias de *Scanning* para iPhone 4S

de ejecución de *scanning* que hace recorrer los canales de forma ascendente (i.e., del canal 1 al 11) y otras que lo hacen en una secuencia descendente (del canal 11 al 1). Luego de un período de tiempo existe una fase de ejecuciones de full scanning que se realizan en secuencias ascendentes. Este comportamiento diferente resulta de la implementación particular de los algoritmos de scanning que el dispositivo Apple iPhone 4S posee y creemos que, buscaría encontrar APs más rápidamente.

B. Sobre la Frecuencia de Ejecución del Proceso Scanning

Los experimentos realizados permitieron generar archivos de captura de traza única que muestran un comportamiento en el que se realiza el proceso de scanning por intervalos de tiempo. En los distintos sistemas operativos se observa que los algoritmos de scanning envían tramas Probe Request en distintos canales por un tiempo y luego esperan la recepción de una trama Probe Response. En la Tabla III se muestran las frecuencias de ejecución observadas para 50 repeticiones por

Table III: Frecuencia de Ejecución de Scanning en Dispositivos Evaluados

Sistema Operativo	Frecuencia de ejecución (s)
Meego 1.2 Harmattan	10
Android 3.2	10
Debian GNU/Linux 6.0	$t_0 = 0$ $t_1 = 20$ $t_2 = t_1 + 10 = 20 + 10 = 30$ $t_i = t_{i-1} + 10$ con $i = 2, 3, 4, 5$ $t_i = 60$ con $i = 6, \dots, n$
Microsoft Windows 7 Ultimate	60
iOS 6.1.2	Fase inicial: no identificable Fase estable: cada 8 s

Table IV: Tiempos de Full Scanning para los Sistemas Operativos de los Dispositivos Evaluados

Sistema Operativo	Duración full scanning (ms)
Meego 1.2 Harmattan	$\mu = 184,35719$ y $\sigma = 30,06954$
Android 3.2	$\mu = 601,59232$ y $\sigma = 11,8229$
Debian GNU/Linux 6.0	$\mu = 113,2495$ y $\sigma = 14,0071$
Microsoft Windows 7 Ultimate	$\mu = 168,0186$ y $\sigma = 22,7389$
iOS 6.1.2	$\mu = 429,23747$ y $\sigma = 25,53178$

experimento.

Para los sistemas operativos Meego y Android se pudo observar valores constantes y cercanos a los 10 segundos. En el caso del sistema Linux el comportamiento es incremental desde el inicio del scanning hasta alcanzar los 60 segundos entre ejecución. El scanning comienza en el tiempo $t_0 = 0s$, luego transcurren 20s para la segunda ejecución en t_1 . A partir de este momento, la próxima ejecución de scanning se realiza en $t_i = t_{i-1} + 10$ con $i = 2, 3, 4, 5$. Para $i = 6, \dots, n$ el tiempo entre ejecución es de 60 segundos.

Para el sistema operativo Microsoft Windows 7 Ultimate se observaron valores constantes de 60 segundos entre ejecución de full scanning. Para el caso del sistema operativo iOS 6.1.2, se observó que la ejecución de full scanning se realiza por fases. Una fase inicial en la que el tiempo aproximado de ejecución entre full scanning no pudo ser determinado. Luego, una fase estable (o predecible) en la cual el tiempo entre ejecución de full scanning se encuentra alrededor de los 8 segundos.

C. Tiempo de Duración del Proceso de Scanning

Castignani, Arcia-Moret y Montavont [1] presentan un estudio del proceso de descubrimiento en redes 802.11. Una de las métricas seleccionadas para caracterizar a este proceso es el tiempo de duración de un full scanning que, corresponde al tiempo transcurrido durante el proceso de descubrimiento para escanear todos los canales disponibles en cualquier orden. En la figura 10 se muestra un bosquejo de la duración del proceso de full scanning.

En la Tabla IV se muestran los tiempos de duración promedio de full scanning para los sistemas operativos de los dispositivos evaluados para 50 repeticiones por experimento. Estos tiempos se encuentran entre los valores sugeridos que varían entre 70ms y 600ms para el retardo de scanning tal como lo reportan Murray, Dixon y Koziniec [23].

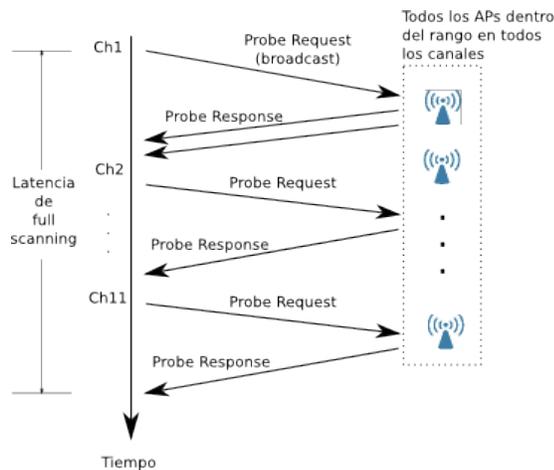


Figure 10: Bosquejo de la Duración de un Full Scanning

VI. CONCLUSIONES Y TRABAJO FUTURO

El prototipo experimental de sniffer para redes 802.11 descrito permitió capturar tramas de administración del estándar IEEE 802.11 en los 11 canales de la banda 2.4GHz. En la construcción se utilizaron componentes de bajo costo como: adaptadores inalámbricos USB, concentradores USB y un computador portátil. El proceso de captura de tramas se automatizó a través de un enfoque modular, cuyo objetivo final es generar una traza única de la actividad de scanning para un dispositivo móvil cualquiera.

Se observó que los dispositivos inalámbricos evaluados presentan distintas secuencias de ejecución de scanning. Unos dispositivos tienen una tendencia a ejecutar el scanning de forma ascendente desde los primeros canales a los últimos de la banda 2.4GHz. Sin embargo, encontramos que pueden haber casos donde se ejecutan el scanning en secuencias alternadas desde los primeros canales a los últimos y viceversa.

La frecuencia de ejecución del scanning es distinta para cada dispositivo. Se observaron valores constantes para su ejecución mientras que otros tienen un comportamiento incremental hasta un umbral como se ve en el sistema operativo Linux.

Sobre la base de los resultados presentados aún no se ha identificado la secuencia de scanning real (la lista de canales propiamente dicha) para los dispositivos evaluados en estricto orden cronológico. El problema de identificación de la secuencia real de scanning puede ser abordado con estudios estadísticos y aplicación de algoritmos más elaborados para identificación de patrones.

En futuras versiones de este trabajo, quisiéramos estudiar la conducta de los algoritmos de scanning en presencia de tráfico de fondo. También planeamos realizar medidas del consumo de energía que aporta un algoritmo de scanning.

AGRADECIMIENTO

Los autores agradecen al Centro de Desarrollo Científico, Humanístico y de las Artes (CDCHTA) de la Universidad de Los Andes por el financiamiento otorgado a través del proyecto I-1369-13-02-B.

REFERENCES

- [1] G. Castignani, A. Arcia-Moret, and N. Montavont, *A Study of Discovery Process in 802.11 Networks*, ACM SIGMOBILE Mobile Computing and Communications Review, New York, USA, 2011.
- [2] Kim, I and Y. Kim, *Prediction-Based Smart Channel Scanning with Minimized Service Disruption for IEEE 802.11e*, IEEE International Conference on Consumer Electronic, Las Vegas, USA, 2011.
- [3] A. Mishra, M. Shin and W.Arbaugh, *An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process*, Process, SIGCOMM Comput. Commun, vol. 33, no. 2, pp. 93–102, New York, USA, 2003.
- [4] *Sniffer de Paquetes y Sistema de Detección de Intrusos para Redes Inalámbricas 802.11*, <http://www.kismetwireless.net>.
- [5] V. Gupta, R. Beyah, and C. Corbett, *A Characterization of Wireless NIC Active Scanning Algorithms*, Wireless Communications and Networking Conference, WCNC 2007, IEEE.
- [6] R. Price and K. Heck, *Performance Comparison of Handover Mechanisms in Wireless LAN Network*, Inst. Comp. Sci., Würzburg Univ, Würzburg, Tech. Rep. 339, December, 2004.
- [7] *IEEE. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE 802.11, 2007*.
- [8] T. Laurensen, *Forensic Data Storage for Wireless Networks: A Compliant Architecture*, M.S. thesis, School of Computing and Mathematical Sciences, AUT Univ., Auckland, 2010.
- [9] C. Corbett, R. Beyah, and J. Copeland, *Using Active Scanning to Identify Wireless NICs*, IEEE Information Assurance Workshop, 2006.
- [10] P. Reddy, H. Sharme, and D. Paulraj, *Multi Channel WiFi Sniffer*. 4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008.
- [11] *Modo Monitor de Operación en Redes Inalámbricas*, <http://wireless.kernel.org/en/users/Documentation/modes>.
- [12] *Interfaz de Programación de Aplicaciones para Captura de Paquetes*, <http://www.tepdump.org>.
- [13] *Estándar de Facto para Inyección y Recepción de Tramas 802.11*, <http://www.radiotap.org>.
- [14] *Raspberry Pi Placa Computadora Desarrollada por la Fundación Raspberry Pi*, <http://www.raspberrypi.org>.
- [15] *SD Specifications. Part 1 Physical Layer Simplified Specification*, https://www.sdcard.org/downloads/pls/simplified_specs/part1_410.pdf.
- [16] *Serial ATA. Estándar de Facto para Almacenamiento Interno en PC*, <https://www.sata-io.org>.
- [17] *Universal Serial Bus Specification, Revision 2.0. 2000*, http://www.usb.org/developers/docs/usb_20_040413.zip.
- [18] *Distribución GNU/Linux diseñada para Auditorías y Pruebas de Penetración Relacionadas con la Seguridad Informática en General*, <http://www.backtrack-linux.org>.
- [19] *Suite Aircrack-ng para Romper Claves WEP y WPAPSK de 802.11*, <http://www.aircrackng.org>.
- [20] *Programa Wireshark para Capturar y Analizar Trazas de Tráfico de Red en Sistemas Operativos Linux*, <https://www.wireshark.org/docs/manpages/tshark.html>.
- [21] *Programa crontab para Administración de Procesos en Segundo Plano de UNIX/Linux*, <http://pubs.opengroup.org/onlinepubs/9699919799/utilities/crontab.html>.
- [22] A. Nicholson, Y. Chawathe, M. Chen, B. Noble and D. Wetherall, *Improved Access Point Selection*, MobiSys'06 Proceedings of the 4th international conference on Mobile systems, applications and services, pp. 233–245, New York, USA, 2006.
- [23] D. Murray, M. Dixon, T. Koziniec, *Scanning Delays in 802.11 Networks*, in proceedings of the 2007 International Conference on Next Generation Mobile Applications, Services and Technologies, pp. 255–260. Washington, DC, USA, 2007.

Gestión de Tráfico Best Effort a través de Retardos Aleatorios en Redes 802.16

Andrés Arcia-Moret^{1,2}, Diego Uzcátegui²
aarcia_m@ictp.it, diego.uzc@gmail.com

¹ International Centre for Theoretical Physics (ICTP), Trieste, Italia

² Grupo de Investigación en Redes y Sistemas Distribuidos (RESIDE), Universidad de Los Andes, Mérida, Venezuela

Resumen: WiMAX es uno de los sistemas de acceso inalámbrico hacia la Internet, que continuamente ha estado evolucionando para prestar mejores servicios de transmisión de datos, voz y video para redes de área metropolitana. En las Redes WiMAX una estación de base (*Base Station* o BS) se encarga de administrar el ancho de banda entre las estaciones suscriptoras (*Subscriber Stations* o SSs) para las distintas clases de servicio. Particularmente, en la asignación de ancho de banda para transmisiones de tráfico *Best Effort* por el canal de subida, desde las SSs a la BS, la BS debe hacer una estimación del ancho de banda requerido por las SSs. Para llevar a cabo este proceso, las SSs envían pequeños paquetes donde solicitan ancho de banda durante un período de contención por trama, que luego son procesados por la BS. En este artículo se presenta un modelo para el procesamiento de las solicitudes de ancho de banda denominado “Gestión del Ancho de Banda con Retardo Aleatorio”, que mejora el rendimiento de redes WiMAX cuando se transmite tráfico TCP. Mediante pruebas de simulación con NS-2 (Modelo WiMAX-NIST), se compara su desempeño con otras políticas y se evalúa en varios escenarios.

Palabras Clave: Redes Inalámbricas; Redes 802.16; Sistema de Solicitud/Otorgamiento de Ancho de Banda; TCP.

Abstract: WiMAX corresponds to one of the Internet wireless access technologies that has continuously evolved offering better data, voice, and video transmissions for metropolitan area networks. In WiMAX networks, a Base Station (BS) manages the bandwidth among the Subscriber Stations (SS) through different Classes of Services. This is a particularly interesting problem for the upload transmissions, i.e., from the SS to the BS, for which the BS has to estimate the required bandwidth to appropriately serve the SS set. During the data transmission process, the SSs send to the BS small control frames for bandwidth request during a contention period, upon which the bandwidth assignment depends. In this article we discuss a model for the processing of these bandwidth requests at the BS called “Bandwidth Management with Random Delay” that improves TCP performance. Using different scenarios, we compare and evaluate with other politics by simulations in NS-2.

Keywords: Wireless Networks; Networks 802.16; Bandwidth Request-Grant Mechanism; TCP.

I. INTRODUCCIÓN

Los sistemas de comunicación digital han evolucionado para cubrir la necesidad creciente de acceso a Internet para los usuarios con difícil acceso, que, por su lejanía de zonas urbanas, no disponen de conexión a la Internet a bajo costo. Uno de estos sistemas es WiMAX, que provee la conexión inalámbrica de última milla a la Internet con una cobertura de Área Metropolitana (*Wireless Metropolitan Access Network* o WMAN), y que promete ser el eslabón siguiente a las redes WiFi. Este calificativo se debe a su capacidad de realizar transmisiones a altas velocidades y con mayor alcance geográfico, entre otras características.

El servicio de red WiMAX provee un esquema de trabajo por conexiones explícitas por cada flujo de datos entre la BS y las

SSs con el fin de soportar mejor los servicios que las distintas aplicaciones requieren. Estas conexiones, determinadas por el identificador de conexión (*Connection Identifier* o CID), son manejadas de diferente forma por las distintas clases de servicio, lo que permite ofrecer un desempeño diferenciado para distintas clases de aplicaciones [1][2][3].

Por otra parte, como es bien conocido, la mayor parte del tráfico que circula en la Internet es transportado por el protocolo TCP (en un 90%). TCP es el protocolo por excelencia para el transporte de datos que, inclusive, hoy por hoy se utiliza en aplicaciones multimedia.

II. REDES 802.16

WiMAX es una tecnología conforme a lo especificado en el estándar IEEE 802.16 [4] y trabaja sobre una topología punto a

multi-punto, donde la Estación Base (*Base Station* o BS) administra el recurso de ancho de banda (AdB) para distribuirlo entre las distintas Estaciones Suscriptoras (*Subscriber Stations* o SSs) (Figura 1). Estas redes transmiten los datos utilizando la técnica de propagación OFDM (*Orthogonal Frequency Division Multiplexing*), lo que ayuda a aprovechar mejor el espectro electromagnético [2].

WiMAX tiene la capacidad de ofrecerle un trato preferencial a flujos específicos de datos, dándole prioridad o reservándole más AdB. Para ello dispone de diferentes niveles de Calidad de Servicio (*Quality of Service* o QoS) para diferentes tipos de aplicaciones. En el estándar IEEE 802.16.a (WiMAX fijo), se contemplan 4 diferentes clases de servicio: UGS (*Unsolicited Grant Service*), donde se provee AdB sin solicitarlo y de forma regular; rtPS (*Real-time Polling Service*), donde se provee AdB bajo solicitud de SS corriendo aplicaciones en tiempo real, en cuyo caso el planificador de la BS respeta ciertas restricciones de tiempo; nrtPS (*Non-real-time Polling Service*), donde a diferencia del anterior las condiciones de cumplimiento del otorgamiento del AdB se relajan un poco, y finalmente, BE (*Best Effort*) donde se transmite el tráfico de la Internet y no se ofrece ningún tipo de garantía. Inclusive, por lo general, en BE se utiliza el AdB que dejan disponible las otras clases de servicio [3][2].

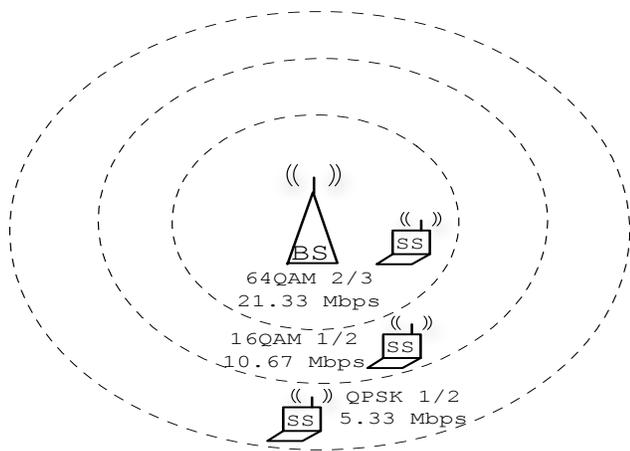


Figura 1: Efecto de los Esquemas de Modulación y Codificación en el Ancho de Banda Bruto Ofrecido en las Redes 802.16

En WiMAX los flujos de datos pueden ser modulados de distintas formas, dependiendo principalmente de la tasa de error permitida para la cobertura deseada. En general, para mayores tasas brutas de transmisión, menores serán las distancias cubiertas. Algunos de los esquemas soportados por el estándar 802.16 son: BPSK (*Binary Phase Shift Keying*), QPSK (*Quadrature Phase Shift Keying*) y QAM (*Quadrature Amplitude Modulation*) [3], así como se muestra en la Figura 1.

Cada trama puede ser individualmente dividida para tráfico bidireccional, sea por tiempo (*Time Division Duplexing* o TDD) o por frecuencia (*Frequency Division Duplexing* o FDD). A su vez se dispone de una subtrama de bajada (*Down-Link* o DL) que corresponde a las transmisiones que van desde la BS hacia las SSs, y de la subtrama de subida que va de las SSs a la BS (*Up-Link* o UL). Como se observa en la Figura 2, la proporción del uso del tiempo o de la frecuencia en cada sub-trama, puede ser adaptada según la necesidad de AdB requerida por cada clase de servicio o en su defecto por el conjunto de flujos. El AdB para la DL es asignado según la

carga de tráfico en el sistema de colas de la BS. En cambio para las UL depende de la eficiencia con que las SSs puedan informar a la BS a través de solicitudes explícitas dentro del sistema de manejo de AdB.

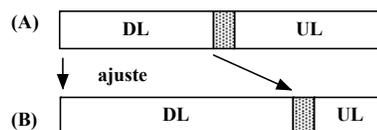


Figura 2: Ajuste de la Capacidad de la Trama WiMAX en Función del Tráfico

A. Gestión del Ancho de Banda en Redes 802.16

La gestión del AdB en Redes 802.16 se refiere más específicamente, a la planificación del uso de este recurso por las SSs. Entonces, el AdB para la UL es asignado a través de lo que en adelante llamaremos, un sistema de gestión colaborativa de ancho de banda o, MSOAB (Mecanismo de Solicitud y Otorgamiento de Ancho de Banda), que consiste en una estimación que hace la BS sobre las necesidades de AdB, luego de procesar pequeños paquetes de solicitud de ancho de banda enviados por las SSs (*Bandwidth Requests* o BW-REQs), con el fin de realizar la asignación del UL de forma adecuada. Estos paquetes son enviados durante un corto período inter-trama accesible bajo un esquema simple de uso del medio; durante cada trama, las estaciones escogen aleatoriamente, de manera uniforme, una celda (o slot) para hacer pasar su respectivo BW-REQ en uno de los canales de contención recomendado en [4].

Las SSs pueden enviar dos tipos de BW-REQs: Incremental, donde el BW-REQ indica la cantidad de AdB que necesitan para transmitir en ese momento, lo que hace que la BS sume esta cantidad a la percepción que está actualmente en la tabla; Agregado, donde en BW-REQ indica la cantidad de AdB total que requieren, lo que hace que la BS sustituya el valor de la tabla por el que indica el BW-REQ.

Para prestar el servicio a la suite de protocolos TCP/IP (para el acceso a la Internet), se usa la clase de servicio BE, que está diseñado para flujos de datos que no requieren un mínimo de AdB y pueden ser tratados en función del AdB disponible.

Según Arcia-Moret y otros [5], existen varias políticas para el manejo de la percepción del AdB que afectan al sistema MSOAB, entre ellas están: RGP (*Reset per Grant Policy*), donde la BS luego de asignar AdB a una SS, coloca en 0 el valor de la percepción de AdB en la tabla, incluso cuando no fue otorgado todo el AdB requerido; DGP (*Decrease per Grant Policy*), donde la BS actualiza la percepción de AdB en la tabla en el momento que el planificador otorga AdB a la SS; y DDA (*Decrease at Data Arrival*), donde la BS actualiza la percepción de AdB en la tabla cuando llegan paquetes de data al servidor. El manejo de los BW-REQs en la política DDA, puede ser dividido a su vez en dos tipos adicionales: iDDA (*Immediate BW-REQ handling DDA*), cuando la BS procesa el BW-REQ una vez que este llega; dDDA (*Delayed BW-REQ handling DDA*), cuando la BS coloca el BW-REQ en una cola, para procesarlo y actualizar la tabla en la siguiente trama. Este último mecanismo advierte un potencial punto de optimización dentro del MSOAB, y es en base a ello que proponemos un nuevo mecanismo en este trabajo.

Arcia y otros [5] encontraron que las transmisiones de datos en Redes WiMAX se pueden mejorar enfocándose en el sistema de manejo de la percepción de las necesidades de AdB para el tráfico transporte. Mediante simulación en NS-2 [6], y utilizando el modelo adoptado por el WiMAX Forum [7], se evaluó el desempeño de políticas para el manejo de los BW-REQs y observo una notable reducción en probabilidad de colisión de los BW-REQs, y una mejora en el rendimiento agregado de los flujos transporte. Algunas de las políticas evaluadas fueron iDDA y dDDA.

Este artículo refleja la continuación de los trabajos de Arcia-Moret y otros [5][8][1][9]; y muestra lo obtenido en Uzcátegui y Arcia-Moret [10]. Aquí presentamos un componente inédito puesto que la gestión de la percepción del ancho de banda en la BS, ha sido muy poco estudiada en la bibliografía. Mucho se ha dicho sobre los planificadores o el cálculo del período de contención [11], pero no sobre la gestión de la percepción. En resumen, este artículo recuenta los puntos relevantes del sistema de gestión de la percepción de consumo de ancho de banda a través de una implementación por simulación. Este estudio es necesario realizarlo por simulación, pues en el mercado no existen equipos WiMAX que permitan acceso al chipset para hacer los estudios propuestos en este artículo.

III. MODELO DE GESTIÓN DE ANCHO DE BANDA CON RETARDO ALEATORIO

En este trabajo, proponemos una extensión a la política dDDA y reportada en [5], descrita en la sección anterior. Esta política la denominamos rDDA (*Random delayed BW-REQ handling - DDA*) y consiste en hacer que la demora expresada en número de trama para procesamiento de los BW-REQs, sea aleatoriamente variable. Para ello, al momento en que llega el BW-REQ a la BS, se genera un número aleatorio λ_i ($1 \leq \lambda_i \leq \tau$) que etiqueta al BW-REQ y que indica la cantidad de tramas que debe esperar éste para ser procesado, es decir, para que la cantidad de AdB solicitada se vea reflejada en la tabla de percepción. La generación del número aleatorio λ_i se obtiene por medio de una distribución de probabilidad uniforme.

La cantidad máxima de tramas que puede esperar un BW-REQ para ser procesado (τ), tiene una estrecha relación con el *timer T16*, el cual está asociado a la espera de la respuesta de la BS a cualquier solicitud hecha a través de un BW-REQ. Cuando T16 expira, indica que la última solicitud de AdB se ha perdido o no ha podido ser atendida. De esta manera queda a potestad de la SS enviar una nueva solicitud para su estado actual de necesidad de AdB.

A. La política rDDA en Marcha

La idea básica de ésta política se muestra en el diagrama de la Figura 3. En T1 la SS solicita AdB a través de un BW-REQ para transmitir 100 bytes. Luego el BW-REQ es recibido por la BS y reservado para ser procesado λ tramas después (en este ejemplo, $\lambda=1$). Dado que se está incorporando una demora en el procesamiento del BW-REQ en ciertos escenarios es posible tener algunas colas de procesamiento vacías. En este caso particular, el algoritmo es capaz de buscar la próxima cola con BW-REQs pendientes por procesar. En T2, transcurridas λ tramas, la BS procesa el BW-REQ, actualiza la tabla de percepción de AdB y otorga los 100 bytes solicitados, para que la SS pueda transmitir en la próxima sub-trama UL. En T3, luego de la aprobación de la BS, la SS puede transmitir 100

bytes de data. Éstos al llegar a la BS, son descontados de la percepción de esa SS reflejada en la tabla. Como una restricción del sistema, para poder realizar otras solicitudes de AdB, la SS debe esperar que se le otorgue el AdB solicitado o que expire el T16 asociado a ese BW-REQs.

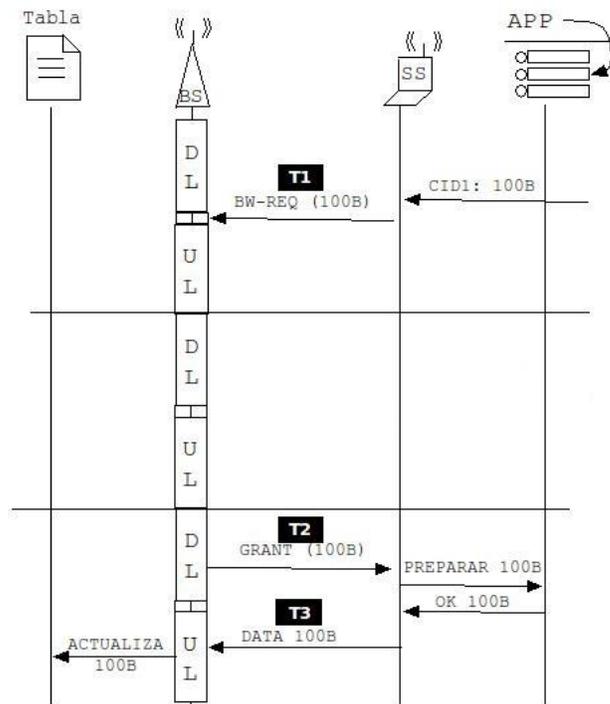


Figura 3: Esquema de Funcionamiento de la Política rDDA

B. Implementación

Esta política se implementó en código fuente C++ y scripts TCL para ser utilizados en NS-2 (*Network Simulator 2*) [6] versión 2.34, que es el simulador más utilizado en el área de las redes. También se utilizó la biblioteca TCP-LAB para facilitar la captura de varias métricas de tráfico TCP. Adicionalmente, se usó el módulo para el MSOAB de WiMAX implementado en [5] usando [7].

Uno de los atributos principales en la implementación de la clase de la nueva política, es una lista circular de colas que contienen las distintas solicitudes o BW-REQs (ver Figura 4). El número de colas es de τ . Cada una de estas colas contiene los BW-REQs que van a ser procesados en una trama futura. También, se usa un indicador a la trama actual para indicar el grupo de BW-REQs que se van a procesar en la trama actual que será donde se construirán los mapas de uso de la trama (DL-MAP). Este cursor cambia, conforme avanza el tiempo y el sistema debe ocuparse de una nueva trama.

El diseño de esta estructura fue concebido con dos premisas. La primera es que, mediante una estructura de lista circular de colas se requiere dar adaptabilidad al sistema para repartir adecuadamente los BW-REQs a lo largo de las tramas para no congestionar en una sola trama ráfagas de solicitudes cuando exista un alza en la demanda, haciendo al sistema escalable. La segunda, es que se introduce la posibilidad clara del manejo de la prioridad en la gestión de BW-REQs en un potencial sistema de colas (por clases de servicios como está concebido WiMAX) mientras que para tráfico BE, se respeta la equidad

entre los flujos a través de un simple algoritmo de repartición aleatoria de los BW-REQ.

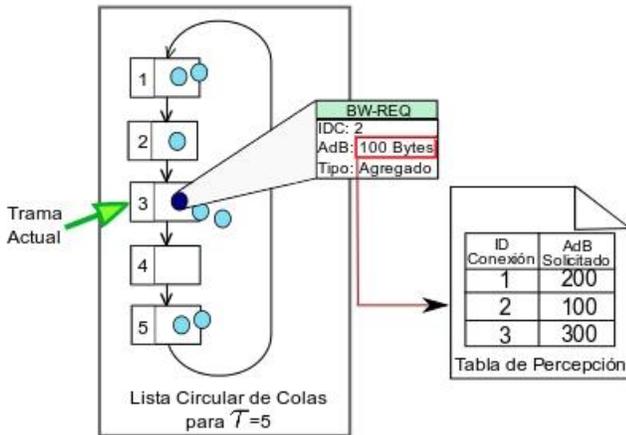


Figura 4: Diseño del Sistema de Lista Circular de Colas para la Política rDDA

IV. PRUEBAS DE FUNCIONAMIENTO

En este estudio se analizó la política rDDA con los valores de $\tau = \{3, 5, 7\}$, con el objetivo de ver el comportamiento cuando se coloca esta política ligeramente diferente a la dDDA (donde el valor τ equivalente sería 1) y para evitar el impacto que pueda tener sobre 'retransmisiones' de BW-REQs debido a expiraciones del valor T16. Se comparó su desempeño con las políticas iDDA y dDDA tanto como para corroborar la correctitud de nuestra implementación, como para hacer comparaciones preliminares con las políticas implementadas en [5]. Se simuló para un número considerable de SSs que varía entre 2 y 32. La topología que se usó en la simulación es la que se muestra en la Figura 5, donde un servidor está conectado a la BS de forma cableada por un enlace de 100 Mbps y un retardo de 2 ms. Luego, las SSs están registradas previamente en la BS para cuando los flujos comienzan a circular. Este escenario corresponde al tipo Data-Center.

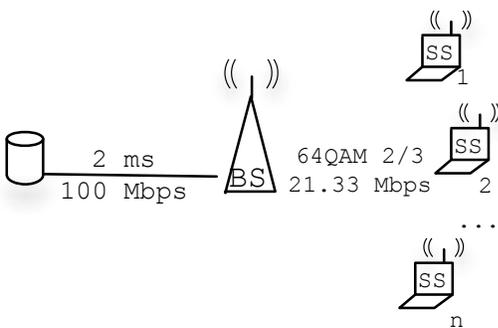


Figura 5: Topología Utilizada para las Pruebas de Simulación

Para cada punto de las gráficas mostradas en la siguiente sección, se realizaron simulaciones de 1000 segundos, usando la modulación 64 QAM $\frac{3}{4}$. La duración de la trama se colocó en 5 ms, y la distribución de las subtramas de bajada y de subida fue 50% del AdB para cada una. Cada SS tuvo una conexión con el servidor donde se transmitieron flujos de datos de una aplicación FTP (en modo descarga o carga de un archivo dependiendo del escenario), la cual está basada en el protocolo TCP. Estos y otros parámetros utilizados en la simulación se resumen en la Tabla I.

Tabla I: Parámetros de la Simulación

Parámetro	Valor
Frecuencia Portadora	5.3 GHz
Ancho de Banda del Canal	7 MHz
Modulación y Codificación	64 QAM $\frac{3}{4}$
Proporción UL:DL	1:1
Tipo de Cola BS	DropTail
Tamaño de la Cola BS	50 paquetes
Duración de la Trama	5 ms
Tipo de BW-REQs	Agregado
Período de Contención	10 slots
--	--
Aplicación	FTP
Versión TCP	Newreno
Tamaño del Segmento TCP	960 Bytes
Esquema de ACKs	Delayed ACK
Duración de la Simulación	1000 segundos

Los escenarios de prueba considerados fueron: Escenario de Descarga, en el que cada una de las SSs descarga un archivo ubicado en el servidor; Escenario de Carga, en el que cada SS coloca un archivo al servidor; y Escenario de Tráfico Cruzado, en el que la mitad de las SSs están descargando un archivo del servidor, y la otra mitad lo están cargando al servidor.

Las métricas utilizadas para la evaluación del rendimiento de la transmisión de datos fueron: *Rendimiento Agregado del Sistema* (Mbps), que corresponde a la suma de todos los paquetes de datos que son entregados por unidad de tiempo; *Probabilidad de Colisión de BW-REQs*, la probabilidad que tiene un BW-REQ de colisionar cuando se envía durante el período de contención inter-trama; *Tasa Agregada de expiraciones de T16*, correspondiente al promedio del total de expiraciones del timer T16 en el sistema por unidad de tiempo; *Índice de justicia de Jain* [12], es una medida popular para el tráfico transporte que indica la equidad en la distribución de los recursos del sistema (en nuestro caso particular es evaluaremos el consumo equitativo de AdB); *Proporción del UL utilizada*, promedio de la relación entre el AdB usado en la sub-trama UL y el AdB disponible en esa sub-trama; *Proporción del DL utilizada*, correspondiente al promedio de la relación entre el AdB usado en la sub-trama DL y el AdB disponible en esa sub-trama; *Promedio de Ocupación de la Cola de Procesamiento de BW-REQs*, cantidad promedio de BW-REQs que están en cola al momento en que se van a procesar, que en promedio puede considerarse también como la cantidad de BW-REQs enviados por trama.

V. RESULTADOS

Para los escenarios de prueba en todos los casos se configuran de 2 a 32 SSs con flujos TCP de larga duración y según lo que se especifica en la Tabla I. Los resultados mostrados en las gráficas corresponden a muestras promediadas luego de 1000 s de simulación, con un previo período de calentamiento donde se espera que la fase inicial de TCP (descubierta del AdB con *Slow Start*) no muestre su impacto. Pues, esta etapa por si sola puede tratarse en un trabajo aparte. La diferencia en las

siguientes subsecciones radica en la dirección de los datos a transmitir.

A. Escenario de Descarga

Este escenario corresponde a los flujos TCP llevando datos desde la BS hasta las SSs. Así como muestra la Figura 4, los clientes SS descargan datos de un servidor en un escenario de Data Center.

En la Figura 6, observamos a modo de corroboración, como la política implementada según la Sección III.A, rDDA (con $\tau = \{3, 5, 7\}$) presenta un rendimiento similar a la política dDDA en todos los casos. Esto además nos muestra que la demora creada para el tratamiento de los BW-REQ no afecta el rendimiento de las conexiones TCP de descarga.

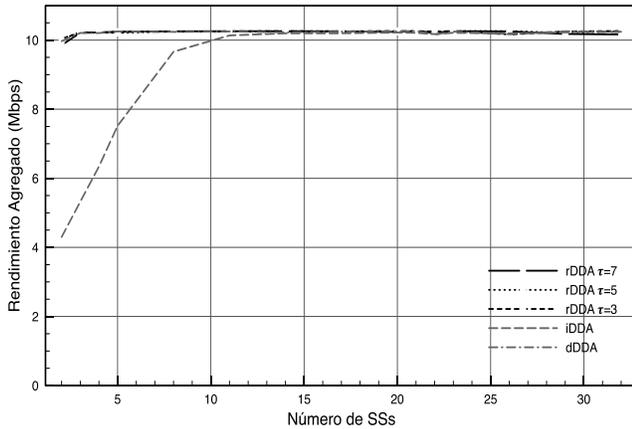


Figura 6: Rendimiento Agregado de las Conexiones TCP en Modo Descarga

En la Figura 7 observamos la tasa agregada de las expiraciones para $T16 = 100$ ms. Recordemos que el *timer T16* expira cuando un BW-REQ no ha sido atendido antes de que transcurran T16 unidades de tiempo o cuando se ha perdido producto de una colisión (que corresponde a la gran mayoría de nuestros escenarios). Observe igualmente que para la política rDDA la tasa de expiración disminuye en un 20% para cargas altas (22 SSs en adelante) que corresponde a la mejor organización de los BW-REQ a lo largo de las colas de tratamiento demorado ($\tau = \{3, 5, 7\}$). Nótese que, producto el mismo efecto es observable al tener una probabilidad de colisión (p) menor para cargas altas donde para $nSS > 22$, p pasa de 0.23 a 0.16.

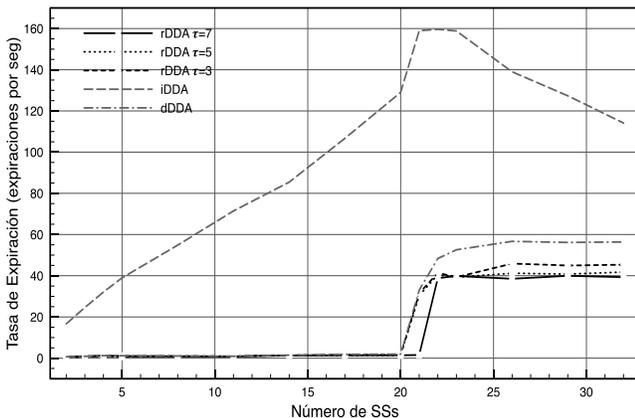


Figura 7: Tasa Agregada de Expiraciones del Timer T16

La Figura 8 muestra la proporción de AdB utilizada en la subtrama de subida. La organización de las solicitudes de AdB producto de la política rDDA en cualquiera de las configuraciones mostradas, hace posible que las subtramas UL puedan ser mejor aprovechadas para cargas superiores a 15 SSs. Esta métrica nos dice que la transmisión de los distintos flujos TCP se hace más suavemente. Esta suavidad se refiere a una transmisión con menos ráfagas, que en otro caso son producto de ACKs acumulados sin ser regularmente servidos (caso de la política dDDA).

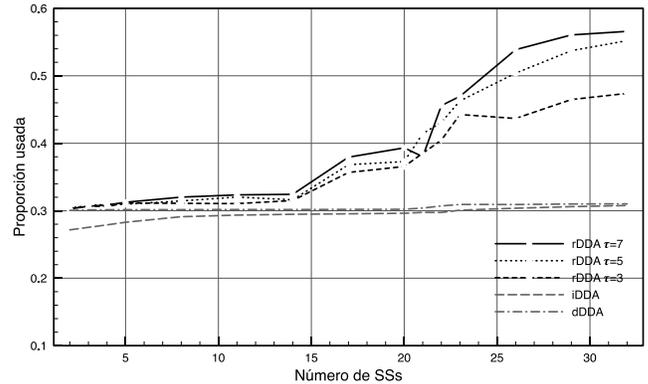


Figura 8: Proporción Utilizada de la Subtrama de Subida (UL)

En la Figura 9 se puede observar que con el uso de la política rDDA, se reduce, sistemáticamente, el envío de BW-REQs en el sistema. Esto, complementado con lo visto en Figura 6, podemos inferir que dicha reducción no afecta el rendimiento del sistema; por lo que resulta beneficioso tener la repartición aleatoria de los BW-REQ en la BS. El procesamiento más rápido de los BW-REQs es debido a que se atienden mejor las solicitudes, generando menos de ellas por unidad de tiempo.

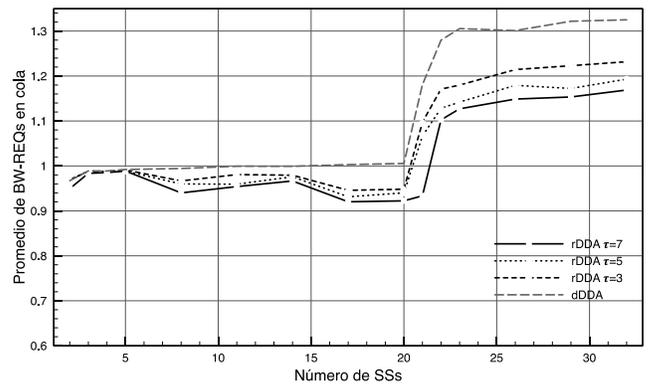


Figura 9: Promedio de BW-REQs en Cola al Momento de Ser Procesados, o de Forma Equivalente, Promedio de BW-REQs Enviados por Trama en la Red

B. Escenario de Carga

Este escenario corresponde a flujos TCP enviados desde las SSs, es decir, los datos van en dirección de la BS y hacia el servidor. En este escenario observamos que la política rDDA obtuvo, en líneas generales, el mismo rendimiento reportado en Arcia y otros [5]. Así, referimos al lector interesado en estos resultados consultar la referencia citada.

C. Escenario de Tráfico Cruzado

Este escenario se contempla transferencias de datos en ambas direcciones (carga y descarga) de forma simultánea. En este

caso, podemos observar el impacto de la mezcla de paquetes de datos y ACKs en una misma trama.

En la Figura 10 observamos el rendimiento agregado para las conexiones de carga y descarga en el escenario de tráfico cruzado. En este escenario apreciamos las bondades de la política rDDA respecto a dDDA, donde se obtiene un incremento en del rendimiento agregado para cargas inferiores a 15 SSs de hasta un 22%. Esta ganancia la atribuimos a lo complementado por la Figura 11, donde podemos ver que la subtrama de subida está casi en su totalidad siendo utilizada durante las transferencias. Es decir, que la subtrama de bajada es subutilizada para menos de 15 estaciones transmitiendo simultáneamente. Sin embargo, nótese que rDDA aprovecha sistemáticamente mejor la repartición del AdB de bajada.

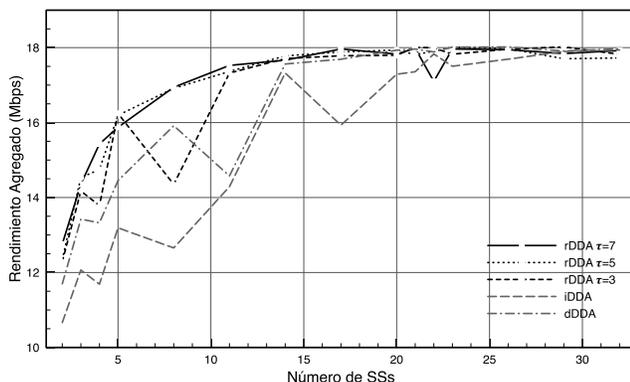


Figura 10: Rendimiento Agregado de Todas las Conexiones en Escenario de Tráfico Cruzado (Carga y Descarga)

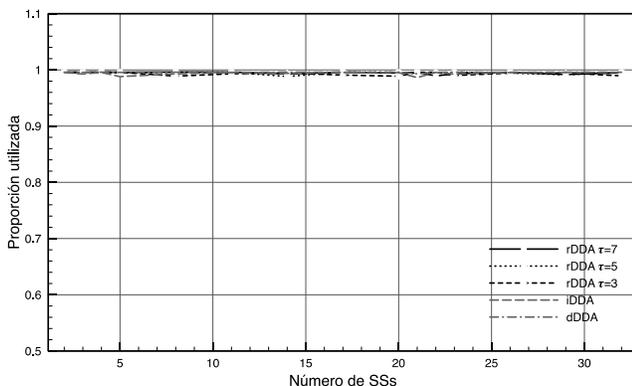


Figura 11: Proporción de Uso de la Subtrama de Subida (UL) en el Escenario de Tráfico Cruzado

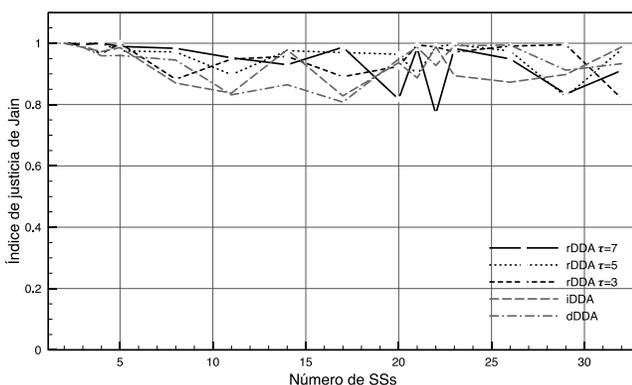


Figura 12: Índice de Justicia de Jain entre los Flujos de Descarga para el Escenario de Tráfico Cruzado

Por último, la Figura 12 muestra que los ACKs compitiendo por recursos en la dirección SS-BS tienen un impacto en el aprovechamiento equitativo del AdB de descarga. Vemos también que el esparcimiento de las solicitudes hechas por rDDA con $\tau = 3$ mantiene un índice de justicia promedio de 0.95, que comparado al índice de 0.9 arrojado por dDDA nos muestra que esparcir las solicitudes y tratarlas a posteriori trae beneficios para el uso equitativo del AdB entre los flujos.

VI. CONCLUSIONES

En este trabajo se diseñó, implementó y evaluó una nueva política llamada “Gestión del Ancho de Banda con Retardo Aleatorio” (rDDA) para la gestión del ancho de banda administrado por la clase de servicio *Best Effort* en Redes 802.16 (WiMAX). Se elaboró exhaustivamente sobre el modelo de tratamiento de las solicitudes de ancho de banda hechas por la estaciones suscriptoras. Al respecto observamos que, como valor añadido a las políticas desarrollada por Arcia y otros [5] en trabajos previos, rDDA es capaz de distribuir más eficientemente el ancho de banda en las descargas, tanto en el caso de flujos exclusivamente en descarga como el caso de tráfico cruzado. Esto representa un beneficio para lo que corresponde al caso común del uso de la Internet, donde los usuarios son consumidores de la información.

La reducción en la tasa de envío de BW-REQs que produce el uso de la política rDDA respecto a dDDA, principalmente cuando hay más de 20 SSs en el escenario de descarga, no afecta en rendimiento del sistema como se puede observar en las respectivas gráficas del escenario de descarga (ver Figura 5 y Figura 8).

Durante la ejecución de este trabajo hemos observado también que el ancho de banda puede ser utilizado mejor aún si prestamos atención al cálculo adecuado del *timer T16* (ver Figura 6, a partir de 20 SS). De igual manera hemos observado que los paquetes transportados en la sub-trama de bajada pueden ser reorganizados mejor para evitar el fenómeno de subutilización del ancho de banda. Estas observaciones son objeto de tratamiento actual para reportes futuros.

REFERENCIAS

- [1] A. Arcia-Moret, Y. Yang, N. Montavon, and D. Ros, *A Study of Bandwidth-Perception Management Mechanisms in IEEE 802.16 Networks*, Technical Report, arXiv.org, Abril 2010.
- [2] A. Arcia-Moret, *Modifications du Mecanisme d'Acquittement du Protocole TCP: Évaluation et Application aux Réseaux Filaires et sans Fils*, PhD Thesis, Telecom Bretagne, Rennes, France 2009.
- [3] C. Cicconetti, L. Lenzini, E. Mingozzi, and C. Eklund, *Quality of Service Support in IEEE 802.16 Networks*, IEEE Network, vol. 20, no. 2, pp. 50–55, 2006.
- [4] IEEE Computer Society and IEEE Microwave Theory/Technique Society, *IEEE 802.16-2009, IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, Agosto 2012.
- [5] A. Arcia-Moret, N. Montavon, A. Paltrinieri, and D. Ros, *Improving Uplink Bandwidth Management in 802.16 Networks*, in proceedings of IEEE Local Computer Networks (LCN), Zurich, Switzerland, October 2009.
- [6] *The Network Simulator - NS-2*, <http://www.isi.edu/nsnam/ns>.
- [7] The National Institute of Standards and Technology (NIST), *The NS-2 WiMAX model*, USA, 2007, http://www.nist.gov/itl/antd/emntg/ssm_tools.cfm.
- [8] A. Paltrinieri, *Improving TCP Performance on 802.16 Networks*, Tesis de Maestría, Universidad de Buenos Aires, 2009.

- [9] Y. Yang, A. Arcia-Moret, N. Montavont, and D. Ros, *On the Impact of IEEE 802.16 Bandwidth Request-Grant Mechanisms on TCP*, in proceedings of Global Communications Conference, Exhibition and Industry (GLOBECOM), December 2010.
- [10] D. Uzcátegui y A. Arcia-Moret, *Diseño, Implementación y Prueba de un Mecanismo de Gestión de Ancho de Banda para Flujos TCP en Redes 802.16*, Tesis, Universidad de Los Andes, Mérida - Venezuela, Mayo 2012.
- [11] J. Delicado, F. Delicado, and L. Orozco-Barbosa, *Study of the IEEE 802.16 Contention Based Request Mechanism*, Telecommunication Systems, vol. 38, no. 1, pp. 19-27, June 2008.
- [12] D-M. Chiu and R. Jain, *Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks*, Computer Networks ISDN Systems, vol. 17, no. 1, pp. 1-14, 1989.

Índice de Autores

A

Acosta, Maribel	34
Alvarado, Ana	34
Álvarez, Silvio	8
Araujo, Antonio	50
Arcia-Moret, Andrés	50, 58

B

Baldizán, Oriana	34
Baranya, Alexander	34

C

Cardoso, Lucia	1
Castillo, Simón	34

D

De Gouveia, Lisset	1
De Simone, Giuseppe	34

G

García, Susana	1
Goncalves, Marlene	34, 43
González, Hancel	34
Guerrero, Argenis	26
Guevara, María	26

H

Herrera, Juan	17
---------------	----

L

La Cruz, Alexandra	34
Losavio, Francisca	17

M

Matteo, Alfredo	17
Mena, Christiam	8
Montoya, Gabriela	34
Moreno, Francisco	8

O

Ojeda, Jordan	8
Ortega, Dinarle	26

P

Paladino, Rosaura	1
Palma, Guillermo	34
Pereira, Wilmer	1

R

Ramírez, Esmitt	8
Rangel, Jeanlight	8
Rodríguez, Omaira	8
Ruckhaus, Edna	34

T

Tineo, Leonid	43
---------------	----

U

Uzcátegui, Diego	58
------------------	----

V

Vidal, Maria-Esther	34
---------------------	----

REVECOM

Sociedad Venezolana de Computación

La Sociedad Venezolana de Computación está comprometida con el impulso de una nueva generación académica y profesional en nuestra área de saber para el desarrollo del país.

Los conceptos y puntos de vista expresados en los trabajos publicados en este libro representan las opiniones personales de los autores y no reflejan el juicio de los editores o de la Sociedad Venezolana de Computación.

ISSN: 2244-7040



9 772244 704006

www.svc.net.ve/revecom

