

**Presentación**  
Presentation

**Editorial: Relación Universidad-Empresa**  
Editorial: Relationships among Universities and Firms  
*Angela María Marulanda-Maya*

**Metaheurísticas Multiobjetivo Cardumen de Peces Artificiales (FAFS) y Optimización Evolucionaria por Enjambre de Partículas con Topología Estocástica Global Individual (FEPSO GIST). Parte II: Aplicación**  
Multiobjective Metaheuristics Artificial Fish School (FAFS) and Evolutionary Particles Swarm Optimization with Global Individual Stochastic Topology (FEPSO GIST). Part II: Application  
*Gustavo Alejandro Schweickardt*

**Análisis de los Indicadores de Emisiones de Gases de Efecto Invernadero para Valorar Proyectos de Energía Renovable en Sistemas Eléctricos, caso de Colombia**  
Analysis Indicators of Greenhouse Gas Emissions for Assessing Renewable Energy Projects in Electrical Systems, Case of Study in Colombia  
*Marco Alejandro Berrío-Monsalve*  
*Mónica Arango-Arango*  
*Sergio Botero-Botero*

**CrawNet: Crawler de Recursos Multimedia para la Web Superficial y Oculta**  
CrawNet: Multimedia Crawler Resources for Both Surface and Hidden Web  
*Alicia Martínez-Rebollar*  
*Fernando Pech-May*  
*Hugo Estrada-Esquivel*  
*Eduardo Pedroza-Landa*

**Paralelización del Algoritmo Expectación-Maximización Utilizando OpenCL**  
Maximization of Expectation-Parallelism Algorithm Using OpenCL  
*Ernesto Insua-Suárez*  
*Marlis Fulgueira-Camilo*  
*Venus Henry-Fuenteseca*

**Análisis de la Penetración de las Tecnologías de la Información y Comunicación (TIC) y su Influencia en la Reducción de la Brecha Digital en el Valle de Aburrá, Caso de Estudio sobre Internet**  
Analysis of the Penetration of Information-Communication Technologies (ICT) and its Influence on Bridging the Digital Gap into the Aburrá Valley's Population, An Internet Case of Study  
*Gloria Amparo Rosales-Acevedo*  
*Sergio Botero-Botero*

**Incorporación de Proactividad a los Agentes en un Observatorio Tecnológico**  
Proactivity Incorporation into Agents of a Technological Observatory  
*Yahima Hadfeg-Fernández*  
*Nelson Morales-Aguilar*  
*Mailyn Moreno-Espino*

**Modelo de Simulación de Estrategias de Inversión para Papicultores Colombianos**  
Simulation Model of Investment Strategies for Colombian Potato Farmers  
*Víctor Daniel Gil-Vera*

**Un Caso de Estudio para la Adopción de un Modelo de Trazabilidad de Requisitos en el Sector Energético**  
A Case Study for the Adoption of a Requirement Tracking Model in the Energy Sector  
*Lina María Montoya-Suárez*  
*Juan Carlos Monsalve-Gómez*  
*Jorge Mauricio Sepúlveda-Castaño*



FUNDACIÓN  
UNIVERSITARIA



Facultad de  
Ingenierías

©Fundación Universitaria Luis Amigó

*Lámpsakos*

N°. 13, Enero-Junio de 2015

**ISSN:** 2145-4086

**Rector**

Pbro. José Wilmar Sánchez Duque

**Vicerrectora de Investigaciones**

Isabel Cristina Puerta Lopera

**Director de Programa Ingeniería de Sistemas**

Ramiro Antonio Giraldo Escobar

**Jefe Departamento de Fondo Editorial**

Carolina Orrego Moscoso

**Diagramación y diseño**

Arbey David Zuluaga Yarce

**Corrector de estilo**

Lillyana Vallejo Gómez

**Traductor**

Gabriel Jaime Correa Henao

**Contacto editorial**

Fundación Universitaria Luis Amigó. Transversal 51A #67 B 90. Medellín, Antioquia, Colombia.

Tel: (574) 448 76 66 (Ext. 9711. Departamento de Fondo Editorial).

[www.funlam.edu.co-fondoeditorial@funlam.edu.co](http://www.funlam.edu.co-fondoeditorial@funlam.edu.co)

Órgano de divulgación de la Facultad de Ingenierías de la Fundación Universitaria Luis Amigó.  
Hecho en Colombia.

Financiación realizada por la Fundación Universitaria Luis Amigó.

**DOI:** <http://dx.doi.org/10.21501/issn.2145-4086>



Licencia Creative Commons Atribución- No Comercial – Compartir Igual 4.0 Internacional.

# LÁMPSAKOS

## Director de la revista

Gabriel Jaime Correa Henao, Ph.D. Fundación Universitaria Luis Amigó. Medellín, Colombia

## Comité científico

Marcelo Becker, Ph.D. Universidade de São Paulo (USP), São Carlos, Brasil

José María Yusta-Loyo, Ph.D. Universidad de Zaragoza (Unizar), España

Gustavo Alejandro Schweickardt, Ph.D. Universidad Tecnológica Nacional, Concepción del Uruguay, Argentina

Bizuayehu Abebe-Worke, Ph.D. Universidade da Beira, Covilhã, Portugal

Aldo Pardo-García, Ph.D. Universidad de Pamplona, Norte de Santander, Colombia

Marta Silvia Tabares-Betancur, Ph.D. Universidad de Medellín (UdeM), Colombia

Yuri Ulianov-López, Ph.D. Universidad Autónoma de Occidente, Cali, Colombia

Andrés Felipe Agudelo-Santamaría, Ph.D. Universidad de Antioquia, Medellín, Colombia

## Comité editorial

Rubén Ángel Galindo-Aires, Ph.D. Universidad Politécnica de Madrid (UPM), España

Giner Alor-Hernández, Ph.D. Instituto Tecnológico de Orizaba, Veracruz, México

Óscar Sapena-Vercher, Ph.D. Universidad Politécnica de Valencia (UPV), España

Amit Chaudhry, Ph.D. Institute of Engineering and Technology, Panjab University, Chandigarh, India

Luis Fernando Garcés-Giraldo, Ph.D. Corporación Universitaria Lasallista, Caldas (Ant), Colombia

Nazly E. Sánchez-Peña, Ph.D. Corporación Universitaria Autónoma del Cauca, Popayán, Colombia

Henry Alonso-Colorado, Ph.D. Universidad de Antioquia, Medellín, Colombia

Novrov Khalidov Dia, Ph.D. Université des Sciences, Technologies et de Médecin, novakchott, Mauritania

Julián Alberto Patiño-Murillo, Ph.D.(c), Universidad Nacional de Colombia, Medellín, Colombia

## Comité de árbitros de la edición

Jorge Alberto Hernández-Cerón, Ph.D. Universidad de Ibagué, Colombia.

Gonzalo Jiménez-Jaramillo, Ph.D. Fundación Universitaria Luis Amigó. Medellín, Colombia.

Paola Andrea Sánchez-Sánchez, Ph.D. Universidad Simón Bolívar, Barranquilla, Colombia.

Franklin Ferraro Gómez, Ph.D. Fundación Universitaria Luis Amigó. Medellín, Colombia

Juan Carlos Rojas-Zerpa, Ph.D. Universidad de los Andes, Mérida, Venezuela.

Julián Alberto Patiño-Murillo, Ph.D(c). Universidad Nacional de Colombia, Medellín, Colombia

Éder Acevedo-Marín, Ph.D(c). Universidad Cooperativa de Colombia, Medellín, Colombia

Juan Carlos Monsalve-Gómez, Ph.D(c). Fundación Universitaria Luis Amigó. Medellín, Colombia.

Víctor Daniel Gil Vera, Ph.D(c). Fundación Universitaria Luis Amigó. Medellín, Colombia

Walter Darién Gómez Torres, MSc. Tecnológico de Antioquia. Medellín, Colombia.

Ángela Isabel Giraldo-Suárez, MSc. Universidad Antonio Nariño. Bogotá, Colombia.

Carlos Alberto Hernández Medina, MSc. Centro Universitario Municipal Camajuaní, Villa Clara, Cuba.

Juan Camilo Giraldo-Mejía, MSc. Universidad de San Buenaventura. Medellín, Colombia

Jorge Mario Gaviria-Hincapié, MSc. Fundación Universitaria Luis Amigó. Medellín, Colombia.

Álvaro Andrés Ramírez Piñeros, MSc. ISAGEN S.A E.S.P. Medellín, Colombia

## Edición

Fundación Universitaria Luis Amigó.

## Solicitud de canje

Biblioteca Vicente Serer Vicens. Fundación Universitaria Luis Amigó. Medellín, Antioquia, Colombia.

## Para sus contribuciones

[lampsakos@funlam.edu.co](mailto:lampsakos@funlam.edu.co)

<http://www.funlam.edu.co/lampsakos>

Facultad de Ingenierías. Fundación Universitaria Luis Amigó. Transversal 51A # 67B-90. Medellín, Antioquia, Colombia.

Tel.: +574 448 76 66, Ext.: 9676

## CONTENIDO

### Presentación

**Editorial: Relación Universidad-Empresa** ..... 10  
*Ángela María Marulanda-Maya*

**Metaheurísticas Multiobjetivo Cardumen de Peces Artificiales (FAFS) y Optimización Evolucionaria por Enjambre de Partículas con Topología Estocástica Global Individual (FEPSO GIST). Parte II: Aplicación** ..... 13  
*Gustavo Alejandro Schweickardt*

**Análisis de los Indicadores de Emisiones de Gases de Efecto Invernadero para Valorar Proyectos de Energía Renovable en Sistemas Eléctricos, caso de Colombia** ..... 24  
*Marco Alejandro Berrío-Monsalve*  
*Mónica Arango-Arango*  
*Sergio Botero-Botero*

**CrawNet: Crawler de Recursos Multimedia para la Web Superficial y Oculta** ..... 39  
*Alicia Martínez-Rebollar*  
*Fernando Pech-May*  
*Hugo Estrada-Esquivel*  
*Eduardo Pedroza-Landa*

**Paralelización del Algoritmo Expectación-Maximización Utilizando OpenCL** ..... 51  
*Ernesto Insua-Suárez*  
*Marlis Fulgueira-Camilo*  
*Venus Henry-Fuenteseca*

## CONTENTS

### Presentation

**Editorial: Relationships among Universities and Firms** ..... 10  
*Ángela María Marulanda-Maya*

**Multiobjective Metaheuristics Artificial Fish School (FAFS) and Evolutionary Particles Swarm Optimization with Global Individual Stochastic Topology (FEPSO GIST). Part II: Application** ..... 13  
*Gustavo Alejandro Schweickardt*

**Analysis Indicators of Greenhouse Gas Emissions for Assessing Renewable Energy Projects in Electrical Systems, Case of Study in Colombia** ..... 24  
*Marco Alejandro Berrío-Monsalve*  
*Mónica Arango-Arango*  
*Sergio Botero-Botero*

**CrawNet: Multimedia Crawler Resources for Both Surface and Hidden Web** ..... 39  
*Alicia Martínez-Rebollar*  
*Fernando Pech-May*  
*Hugo Estrada-Esquivel*  
*Eduardo Pedroza-Landa*

**Maximization of Expectation-Parallelism Algorithm Using OpenCL** ..... 51  
*Ernesto Insua-Suárez*  
*Marlis Fulgueira-Camilo*  
*Venus Henry-Fuenteseca*

<b>Análisis de la Penetración de las Tecnologías de la Información y Comunicación (TIC) y su Influencia en la Reducción de la Brecha Digital en el Valle de Aburrá, Caso de Estudio sobre Internet .....</b>	<b>62</b>
<i>Gloria Amparo Rosales-Acevedo</i>	
<i>Sergio Botero-Botero</i>	
<b>Incorporación de Proactividad a los Agentes en un Observatorio Tecnológico .....</b>	<b>72</b>
<i>Yahima Hadfeg-Fernández</i>	
<i>Nelson Morales-Aguilar</i>	
<i>Mailyn Moreno-Espino</i>	
<b>Modelo de Simulación de Estrategias de Inversión para Papicultores Colombianos .....</b>	<b>81</b>
<i>Víctor Daniel Gil-Vera</i>	
<b>Un Caso de Estudio para la Adopción de un Modelo de Trazabilidad de Requisitos en el Sector Energético .....</b>	<b>88</b>
<i>Lina María Montoya-Suárez</i>	
<i>Juan Carlos Monsalve-Gómez</i>	
<i>Jorge Mauricio Sepúlveda-Castaño</i>	

<b>Analysis of the Penetration of Information-Communication Technologies (ICT) and its Influence on Bridging the Digital Gap into the Aburrá Valley's Population, An Internet Case of Study .....</b>	<b>62</b>
<i>Gloria Amparo Rosales-Acevedo</i>	
<i>Sergio Botero-Botero</i>	
<b>Proactivity Incorporation into Agents of a Technological Observatory .....</b>	<b>72</b>
<i>Yahima Hadfeg-Fernández</i>	
<i>Nelson Morales-Aguilar</i>	
<i>Mailyn Moreno-Espino</i>	
<b>Simulation Model of Investment Strategies for Colombian Potato Farmers .....</b>	<b>81</b>
<i>Víctor Daniel Gil-Vera</i>	
<b>A Case Study for the Adoption of a Requirement Tracking Model in the Energy Sector .....</b>	<b>88</b>
<i>Lina María Montoya-Suárez</i>	
<i>Juan Carlos Monsalve-Gómez</i>	
<i>Jorge Mauricio Sepúlveda-Castaño</i>	

# PRESENTACIÓN

La revista *Lámpsakos* es una publicación adscrita a la Facultad de Ingenierías de la Fundación Universitaria Luis Amigó, Medellín, Colombia. Tiene por objetivo divulgar artículos originales e inéditos como resultados de investigación científica y tecnológica de las áreas de ingenierías. Fundada en 2009, la revista tiene una periodicidad semestral (se publica en junio y diciembre) con alcance nacional e internacional, tiene circulación *online* de libre acceso, bajo estándares de rigor científico y de calidad editorial.

Los temas tratados en la revista se dirigen a todas las personas de los espacios académicos, investigativos, científicos y profesionales de los diferentes sectores educativos, productivos y empresariales que desarrollan sus actividades en torno a la inge-

nería. Los trabajos que publica corresponden a aspectos de desarrollo científico, profesional y tecnológico en áreas como ingeniería informática, sistemas de información, ingeniería de requerimientos, ingeniería electrónica, ingeniería eléctrica, sistemas inteligentes, control, automatización y robótica, ciberseguridad, ingeniería ambiental, gestión de recursos hidráulicos, gestión del agua, eficiencia energética, simulación de sistemas, investigación de operaciones, gestión del conocimiento en ingeniería, educación en ingeniería, gerencia de proyectos, control de procesos, química, desarrollo de materiales, termodinámica, mercados de la energía, física, nanotecnología, estrategias de mantenimiento industrial, modelos computacionales, urbanismo y paisajismo arquitectónico.

# PRESENTATION

*Lámpsakos* journal is a publication ascribed to the Faculty of Engineering and Architecture of Fundación Universitaria Luis Amigó, in the city of Medellín, Colombia. It aims to divulge original and unpublished articles as a result of scientific and technological research in the areas of engineering. It is published twice a year (issued in June and December) with national and international coverage. The journal is a free access online magazine, published under standards of scientific rigor and editorial quality.

Topics covered in the journal are addressed to all persons in academic, researching, scientific and professional areas from different educational and productive sectors and firms whose activities are performed

on engineering. The journal publishes papers of scientific, professional and technological fields such as Computer Science, Information Systems, Software Requirements, Electronics Engineering, Electrical Engineering, Intelligent Systems, Control, Automation and Robotics, Cyber Security, Environmental Engineering, Water Resources Management, Water Management, Energy Efficiency, System Simulation, Operations Research, Engineering Knowledge Management, Education, Engineering, Project Management, Process Control, Chemistry, Materials Development, Thermodynamics, Energy Markets, Physics, Nanotechnology, Industrial Maintenance Strategies, Computational Models, Urbanism and Architecture.

# ORIGINALIDAD Y DIVULGACIÓN DE ARTÍCULOS

## PAPER ORIGINALITY AND ARTICLE RELEASING

Los trabajos publicados en la revista, sin excepciones, se acogen a las normas del Copyleft y Creative Commons, ya que la publicación es de libre distribución para el conocimiento y podrá ser reproducida por cualquier medio, haciendo mención de la fuente.

El autor o autores aceptan las políticas editoriales y los lineamientos de la guía y autorizan a la revista y a la Institución para editar y divulgar/publicar el artículo por cualquier medio nacional y/o internacional, impreso o electrónico.

Los artículos son producto de un proceso de investigación de los autores y han sido valorados previamente por colegas expertos antes de ser presentados a publicación.

Las opiniones y afirmaciones que aparecen en la publicación reflejan exclusivamente los puntos de vista de sus autores y no comprometen necesariamente las políticas y pensamiento de la FUNLAM, la Facultad de Ingenierías y Arquitectura o la revista *Lámpsakos*.

All papers that are published on this media, must gather to Copyleft and Creative Commons policies, with no exceptions, since they belong to an open access academic journal. Contained information may be reproduced on any media by correctly citing the reference source.

The author or authors authorize the magazine and the Institution to edit and disseminate / publish the article by any media national and / or international, print or electronic.

Papers presented in the academic journal are mostly derived from research activities driven by their authors. They have been previously evaluated on a peer review process before concerning their publication. Each manuscript undergoes an originality evaluation before starting the peer-review process.

The opinions and affirmations accepted for publication reflect exclusively the points of view of their authors and do not necessarily compromise the policies and thinking of either Fundación Universitaria Luis Amigó, the Faculty of Engineering and Architecture or the journal "*Lámpsakos*".



# EDITORIAL

## Relación Universidad-Empresa Relationships among Universities and Firms

**Ángela María Marulanda-Maya, Esp.**

*Gerencia General  
Periódico El Mundo  
Medellín, Colombia  
amarulanda@elmundo.com*

(Recibido el 02-12-2014. Aprobado el 20-12-2014)

Desde todos los puntos de vista, la relación Empresa-Universidad constituye un enriquecimiento para todos los involucrados. El estudiante, al acercarse a la cultura empresarial, vive y siente el ambiente, la cultura que se crea y construye en las empresas todos los días; sin importar si esta es de carácter privado o público.

En este intercambio aparecen factores diferenciadores, dependiendo de la actividad de la empresa, su filosofía y proyecciones de negocio. Desde la más pequeña o la más grande, todas aportan y brindan nuevas experiencias a las cuales los estudiantes deben estar abiertos y dispuestos todo el tiempo como esponjas que luego decantan y se quedan con lo que más les convenga en su formación personal y profesional.

La cultura de la empresa es muy diferente a la academia, pero sin la formación académica, no podríamos llegar a ser y tener el nivel requerido y exigido en las empresas. Independiente del tipo de Ingeniería que escojamos estudiar, es muy importante esforzarnos por ser los mejores cada día, buscando estar actualizados en tendencias, conceptos y todo lo que consideremos debamos aprender desde la parte profesional para cumplir a cabalidad nuestras tareas y proyectos.

Hay momentos en los cuales debemos afianzar y ampliar nuestra formación académica y no debemos escatimar en hacerlo, y es allí donde la academia nos abre las puertas para que cada día nos formemos.

Todas las empresas tenemos un papel importante de apoyo a la academia al servir como primera experiencia en la formación de los estudiantes; debemos recordar que todos pasamos por ahí y sin ese primer acercamiento, no hubiéramos entendido lo diferente que es trabajar y estudiar, y que la realidad de las empresas se complementa perfecto con lo aprendido, siempre y cuando apliquemos como principio fundamental el sentido común y la responsabilidad, sin importar qué experiencia tenemos o no.

En el Periódico El Mundo tenemos claro que el aprendizaje empresarial es fundamental y, en las diferentes áreas de la empresa, nos esforzamos para que nuestros practicantes conozcan, aprendan y asuman sus responsabilidades considerando los límites de experiencia que existen frente a las tareas encomendadas.

Como empresa, todos debemos tener claro que los estudiantes están en un proceso de acercamiento al mundo profesional y no por eso debemos abusar en horarios o tareas encomendadas.

# ARTÍCULOS DE INVESTIGACIÓN



# Metaheurísticas Multiobjetivo Cardumen de Peces Artificiales (FAFS) y Optimización Evolucionaria por Enjambre de Partículas con Topología Estocástica Global Individual (FEPSO GIST). Parte II: Aplicación

Multiobjective Metaheuristics Artificial Fish School (FAFS) and Evolutionary Particles Swarm Optimization with Global Individual Stochastic Topology (FEPSO GIST). Part II: Application

**Gustavo Alejandro Schweickardt, PhD.**

*Grupo de Investigación sobre Economía Computacional de Regulación de Redes y Energías Renovables  
Universidad Tecnológica Nacional  
Concepción del Uruguay, Argentina  
gustavoschweickardt@conicet.gov.ar*

(Recibido el 20-10-2014. Aprobado el 20-12-2014)

**Resumen.** El presente trabajo describe un modelo para la Optimización del Grado de Desbalance de Cargas en una Red Trifásica de Distribución de Energía Eléctrica (SDEE) en Baja Tensión (BT). Se presenta la integración de dos novedosas metaheurísticas: la FEPSO GIST (*Fuzzy-MultiObjective Particles Swarm Optimization with Global/Individual Stochastic Topology*), desarrollada por el autor, y la FAFS (*Fuzzy-MultiObjective Artificial Fish School*), cuya extensión multiobjetivo es propuesta por el autor, valuando la función de aptitud mediante conjuntos difusos. El problema propuesto ya ha sido resuelto en un trabajo presentado por el autor, mediante la metaheurística FPSO, y su solución constituye una referencia para comparar resultados. Entre los inconvenientes producidos por un elevado grado de desbalance en las fases del sistema, se consideran la minimización de las pérdidas técnicas y la mejo-

ra del perfil de tensiones. Ambos aspectos, relativos al uso racional de la energía propiciado desde el lado de la oferta, son observados por la autoridad regulatoria. Se presenta, adicionalmente, un modelo matheurístico que combina el enfoque clásico del problema, empleando programación lineal entera-mixta con las dos metaheurísticas introducidas, FEPSO GIST y FAFS. Se comparan resultados al aplicar los modelos sobre el mismo SDEE BT considerado en el enfoque resuelto vía FPSO. En esta Parte II, se desarrollan los modelos específicos y se presenta su aplicación al problema.

**Palabras clave:** Cardumen de peces artificiales; desbalance de cargas; enjambre de partículas; metaheurísticas; sistemas de distribución de energía eléctrica.

**Citación de artículo, estilo IEEE:**

G.A. Schweickardt, "Metaheurísticas Multiobjetivo Cardumen de Peces Artificiales (FAFS) y Optimización Evolucionaria por Enjambre de Partículas con Topología Estocástica Global Individual (FEPSO GIST). Parte II: Aplicación", *Lámpsakos*, N° 13, pp. 13-23, 2015.

**DOI:** <http://dx.doi.org/10.21501/issn.2145-4086>

**Abstract.** This work presents a model to Low Voltage (LV) Unbalance Degree Optimization in a Three-phase Electric Distribution Network (EDN). The combination of two new Fuzzy-MultiObjective MetaHeuristics FEP SO GIST (*Fuzzy Particles Swarm Optimization with Global/Individual Stochastic Topology*) proposed by the author and, FAFS (*Fuzzy Artificial Fish School*) extended to MultiObjective domain by the author, using Fuzzy Sets, are presented. Of multiple problems resulting from such unbalance degree, are considered the technical losses and the voltage drops. Both aspects are fundamentals in the rational use of energy, when this objective is focused from the

offer side, and are observed for Regulatory Authority. In addition, a MatHeuristic approach composed for the classical approach based in Mixed-Integer Linear Programming and FEP SO GIST-FAFS MetaHeuristics, is introduced. In this Part II of the work, the Models applications to the Unbalance Load Degree problem and its specifics developments, are presented.

**Keywords:** Artificial Fish School; electric distribution systems; metaheuristics; particles swarm; load unbalance.

## 1. INTRODUCCIÓN

En la primera parte del presente trabajo, se dejó definido el problema de balance de cargas, mediante una formulación cualitativa que se corresponde con un problema de optimización multiobjetivo. Asimismo, se presentaron dos enfoques de solución clásicos: a) uno menos riguroso, desde el punto de vista metodológico, soportado en una búsqueda iterativa (BI), que recurre a heurísticas, en el caso más general, diseñadas conforme el conocimiento del SDEE BT específico bajo estudio, y b) otro formal, planteado como un modelo de programación lineal entera-mixta, el cual introduce como hipótesis una característica de las cargas de Intensidad constante (pues, en otro caso, el problema no podría ser linealizable). Su resultado, dada esta consideración muy alejada de la realidad, fue de inferior calidad que el obtenido mediante el enfoque BI, el cual, al menos, asume una característica de las cargas de potencia constante.

Pudo justificarse entonces el empleo de algoritmos de optimización bio-inspirados, tales como algunas metaheurísticas cuyos diseños se sustentan en la Inteligencia de grupo. Por ello fueron explicados los cinco principios de la inteligencia de grupo (IG) y presentados los desarrollos de las metaheurísticas designadas como X-PSO formas (variantes de la *Particle Swarm Optimization* u Optimización por Enjambre de Partículas) y la AFS (*Artificial Fish School* o Cardumen de Peces Artificiales). Tales metaheurísticas son aplicables a problemas de optimización mono-objetivo. De modo que en esta segunda parte del trabajo, se desarrollarán los aspectos teóricos específicos que conducen a la formulación de la forma X-PSO multiobjetivo, referida como FEPSO GIST [1], así como el algoritmo AFS y su extensión al dominio multiobjetivo, referido como FAFS. Ambas se compararán y se integrarán a los efectos de componer la mejor estrategia de solución para el problema de balance de cargas. Esta segunda parte del trabajo está organizada en la forma siguiente: En la sección 2, se presentan específicamente las metaheurísticas EPSO GIST y AFS, ambas primigeniamente aplicables al dominio mono-objetivo en la solución de problemas de optimización. Se describe, considerando las características del espacio de soluciones de la clase de problemas a la que pertenece el balance de cargas en un SDEE BT (espacio discreto de buenas soluciones muy aisladas), el comportamiento de la X-PSO forma EPSO GIST en términos de la satisfacción de los principios de la inteligencia de grupo,

sacando provecho de su topología de comunicación entre partículas, para cada instancia de búsqueda. También se describe la ventaja de integración de las metaheurísticas EPSO GIST y AFS, a los efectos de complementar sus capacidades individuales de satisfacción de los principios de IG. En la Sección 3, se desarrollan los conceptos y se presenta el enfoque propuesto para extender ambas metaheurísticas al dominio multiobjetivo (en rigor, válido para cualquiera de las X-PSO formas presentadas en la primera parte). Para ello se apela a la *toma de decisión en ambientes difusos*, empleando conjuntos difusos en la representación tanto de los objetivos como de las restricciones contenidos en la formulación *rigida* de un problema de optimización multiobjetivo. Surgen de esta extensión tres requerimientos más adicionales a la satisfacción de los cinco principios IG: a) la capacidad para captar las incertidumbres de valor en las variables solidarias a los objetivos/restricciones del problema; b) La aquí introducida como compatibilidad métrica del operador seleccionado para evaluar la aptitud de una solución en el espacio homónimo; y c) la Pareto-dominancia entre un conjunto de buenas soluciones igualmente satisfactorias. En la sección 4, se presenta la formulación de los conjuntos difusos para cada objetivo del problema de balance de cargas. Asimismo, se sugieren diferentes estrategias de solución, hasta arribar a una metaheurística que integra la modelación MIP y una combinación de las metaheurísticas FEPSO GIST + FAFSA, donde la letra “F” corresponde a las versiones multiobjetivo para las mismas (F = *Fuzzy* o Difuso). En la Sección 5, se presentan las simulaciones y resultados obtenidos.

Finalmente, se plantea una de las posibles líneas de investigación futuras (trabajos futuros) y se presentan las conclusiones. Se han incorporado las REFERENCIAS [1, 5–10], citadas en la Parte I del trabajo completo.

## 2. METAHEURÍSTICAS EPSO GIST Y AFS

### 2.1. La metaheurística EPSO GIST

Esta propuesta de variante X-PSO presenta una alteración respecto de la topología de comunicación entre las partículas en cada instancia de decisión o iteración algorítmica respecto de la que emplea la X forma EPSO. Por topología de comunicación debe entenderse la manera en que las partículas de la población son informadas de la localización del óptimo

global en cierta instancia de decisión (iteración algorítmica)  $k$ -ésima. Desde el planteo y diseño de la ecuación de movimiento PSO, se asume que la topología es tipo estrella global (*Global Star Topology*), lo cual significa, virtualmente, que la/las partícula/s que ocupa/n la posición correspondiente al óptimo global, en cierta instancia de decisión, se comportan como el centro de una estrella que “irradia” la información relativa a tal posición hacia el resto de las partículas de la población o enjambre. Luego de varias investigaciones por parte del autor, en co-autoría con otros investigadores, sobre la componente de *cooperación o comunicación* en la ecuación de movimiento PSO con esta topología, en espacios de soluciones cuyas características resultan las del problema aquí abordado, se observó una inevitable tendencia hacia una *convergencia prematura* del algoritmo metaheurístico [2]. Por *convergencia prematura* debe interpretarse una condición en la que el enjambre de partículas queda “atrapado” en soluciones de escasa calidad (frecuentemente referidas, preservando la terminología de la optimización clásica, como óptimos locales, en contraposición al óptimo global buscado). De manera que los cinco principios IG resultaban gradualmente y, conforme aumentaba el número de iteraciones, cada vez más insatisfechos.

Por tanto, se intentó resolver esta limitación desarrollando, esencialmente, las X-PSO formas diferentes de la Canónica, presentadas en la primera parte de este trabajo, logrando mejoras destacables, que minimizaban notoriamente el auto-ajuste de los parámetros de la ecuación de movimiento, al concebir la forma EPSO mediante los operadores evolucionarios. Pero la topología de comunicación se sostenía en su idea inicial: Estrella global y en espacios de soluciones caracterizados como se dijo; si bien con menos frecuencia, la *convergencia prematura* persistía, dependiendo del problema a resolver. De modo que se propuso *alterar* de algún modo la topología de comunicación. Se pensó, entonces, modificarla con un criterio tal que su nueva estructura propendiese a la *diversificación de las soluciones*. Esto significa mejorar el principio de IG indicado como IG3 en la primera parte del presente trabajo, la diversidad de respuesta, promoviendo la posibilidad de que las partículas del enjambre *tengan respuestas diferentes ante los mismos estímulos*. De este modo, existiría una respuesta diferente (una dirección de movimiento diferente en el espacio de soluciones) definida desde la instancia de decisión  $k$  hacia la  $k+1$ , *para ciertas partículas*, permitiendo que continúen “buscando” una *solución de mejor calidad*, ante un estímulo que,

bajo la topología estrella global, conduciría al enjambre a la condición de *convergencia prematura* (en términos de la optimización clásica, sería así posible “sortear” el óptimo local). Siguiendo tal criterio, se propuso la X-forma EPSO GST, cuya topología de comunicación es estrella, lo cual, y conforme a lo explicado, se dará, de aquí en más, por implícito, global estocástica (GST = *Global Stochastic Topology*). Esto fue implementado del siguiente modo: la información relativa al óptimo global, OG, tiene asignada una probabilidad de ser comunicada al resto de las partículas del enjambre, la cual viene dada por una variable aleatoria con distribución gaussiana de media 0 y varianza 1,  $N(0,1)$ .

De este modo, si  $p_i$  es la probabilidad de que la partícula  $i$ -ésima del enjambre reciba información sobre la posición OG en cierta instancia de decisión  $k$ , puede ocurrir que: a) si  $p_i = 1$ , la partícula se comporte como si la topología fuese global; b) si  $p_i = 0$ , la partícula no reciba información sobre el OG y el término de la ecuación de movimiento correspondiente a la cooperación, no tenga efecto; o c) si  $0 < p_i < 1$ , tal término impacte proporcionalmente al valor de  $p_i$ . Así, resulta que la ecuación evolutiva de movimiento u operador velocidad para la forma EPSO GST, vendrá dada por:

$$\mathbf{V}_i^{[k+1]} = w_{il}^{[k]} \times (\mathbf{r}_1^{[k]}) \times \mathbf{V}_i^{[k]} + w_{ic}^{*[k]} \times (\mathbf{r}_2^{[k]}) \times [\mathbf{b}_i^{[k]} - \mathbf{X}^{[k]}] + w_{is}^{*[k]} \times (\mathbf{r}_3^{[k]}) \times [\mathbf{b}_G^{[k]} - \mathbf{X}^{[k]}] \times \alpha \text{OG}^{[k]} \quad (1)$$

- En (1) todos los parámetros tienen la significación dada al describir la X-PSO EPSO, en la primera parte del presente trabajo,
- $\alpha \text{OG}^{[k]i}$  es la probabilidad de que el óptimo global (OG) afecte al movimiento de la partícula  $i$ , en la iteración  $k$ . Nótese cómo el tercer término, correspondiente a la *cooperación o comunicación* entre las partículas del enjambre, se ve afectado en la forma dicha, al incorporar la probabilidad  $\alpha \text{OG}^{[k]i}$  como factor.

Concebida esta metaheurística EPSO GST, pudo comprobarse una sensible mejora respecto de la *convergencia prematura* referida, preservando la mayor capacidad de auto-ajuste de parámetros de la EPSO.

Se la aplicó individualmente, optimizando cada uno de los objetivos que fueron definidos a presentar el problema de balance de Carga en la primera parte del presente trabajo: a) Minimización de pérdidas de potencia activa; b) Minimización de la componente



homopolar a la salida del alimentador principal del SDEE BT analizado; y c) Maximización del índice de tensión de suministro en cada nodo. En este último problema, se pretende, habida cuenta de la operación radial del SDEE BT, que todos los nodos terminales en los alimentadores derivados tengan una tensión que no se aparte, en defecto, más de un 5% respecto de la tensión nominal. Esto lo fija, como se dijo en la primera parte del presente trabajo, la Autoridad Regulatoria Argentina; es decir que, modificando las conexiones a las distintas fases del sistema trifásico de las acometidas domiciliarias monofásicas, debería arribarse a esa condición respecto de la tensión de suministro. El espacio de soluciones para este objetivo en particular exhibe soluciones mucho más aisladas que en el resto de los objetivos, al punto tal que, aún empleando la metaheurística EPSO GST, es necesario, dependiendo del tamaño (número de cargas a reconectar) del problema, emplear una técnica de inicialización, llamada de *arranque o bootstrapping*, para que se localice alguna condición, desde la cual la EPSO GST comience a mejorar las soluciones. Caso contrario, no puede evitarse la *convergencia prematura* del algoritmo.

De manera que se propone enfatizar la satisfacción del principio IG3, aumentando la diversidad de respuesta de las partículas ante el mismo estímulo, modificando, nuevamente, la topología de comunicación. Para ello, el criterio empleado estriba en una extensión del que se utilizó para desarrollar la GST, introduciendo una aleatoriedad mayor. Esta se sustenta en no solo limitar la información que reciben las partículas sobre la posición OG en cierta instancia de decisión k, sino en consignar una probabilidad de que cada partícula reciba información sobre el óptimo individual (OI) de las demás. Con este criterio, existiría una componente de *cooperación o comunicación* adicional, sustentada en la *memoria autobiográfica*, pues *todas o algunas* partículas estarían comunicando a las demás la información de su OI alcanzado en la instancia de decisión k.

Operacionalmente, la probabilidad referida también viene dada por una *variable aleatoria* con *distribución gaussiana* de media 0 y varianza 1,  $N(0,1)$ .

Al involucrarse tanto la Información del OG como del OI de cada partícula, cuya incidencia en la dirección del movimiento para la partícula i-ésima depende de una probabilidad, la topología de comunicación

resultante se denomina GIST (*Global/Individual Stochastic Topology* – Topología Estocástica Global/Individual).

De modo que, aplicada sobre la X forma EPSO, se concibe una nueva X-PSO, denominada EPSO GIST.

La ecuación evolutiva de movimiento u operador velocidad para la forma EPSO GIST, vendrá dada por:

$$\mathbf{V}_i^{[k+1]} = w_{il}^{*[k]} \times (\mathbf{r}_1^{[k]}) \times \mathbf{V}_i^{[k]} + w_{ic}^{*[k]} \times (\mathbf{r}_2^{[k]}) \times [\mathbf{b}_i^{[k]} - \mathbf{X}_i^{[k]}] + w_{is}^{*[k]} \times (\mathbf{r}_3^{[k]}) \times [\mathbf{b}_i^{[k]} - \mathbf{X}_i^{[k]}] \times \alpha \text{OG}_i^{[k]} + \sum_{j=1, j \neq i}^n (\delta(k) \times w_{ic}^{*[k]} \times [\mathbf{b}_j^{[k]} - \mathbf{X}_j^{[k]}] \times \alpha \text{OI}_{ij}^{[k]}) \quad (2)$$

Donde

- $\text{TOI}_{ij}^{[k]} = \sum_{j=1, j \neq i}^n (\delta(k) \times w_{ic}^{*[k]} \times [\mathbf{b}_j^{[k]} - \mathbf{X}_j^{[k]}] \times \alpha \text{OI}_{ij}^{[k]})$  : Resultan ser los nuevos términos de *cooperación* entre la partícula i-ésima y las del resto del enjambre, j-ésimas, con  $j \neq i$  y  $j$  en  $[1..n]$ , siendo n el número de partículas del enjambre.
- $\text{OI}_{ij}^{[k]}$  es la probabilidad de que el óptimo Individual (OI) de la partícula j afecte al movimiento de la partícula i,  $j \neq i$
- $\delta(k)$  es un *factor multiplicativo* o de *decaimiento inercial*, que va haciendo “decaer” la influencia de los OI, con probabilidades distintas de 0, desde cada partícula del enjambre hacia el resto, conforme aumenta el número de iteraciones.  $\delta(k)$  se ha introducido para limitar, en ciertos casos observados, la influencia excesiva del principio IG3, diversidad de respuesta, afectando al principio IG4, estabilidad. Su construcción sigue la misma estructura propuesta para la X-PSO con función de decaimiento inercial, en la primera parte del presente trabajo.

En el caso límite para cierta instancia de decisión k pueden tener que computarse en (2), n-1 términos  $\text{TOI}_{ij}^{[k]}$ . Siendo el número de partículas usual para esta clase de problemas  $n = 200$ , caso típico, tal cálculo no supone ninguna complejidad de cómputo adicional si se consideran los beneficios que la EPSO GIST aporta a los efectos de evitar las *convergencias prematuras* en situaciones tal como la explicada respecto de la tensión de suministro para el problema de balance de cargas.

Al abordar la extensión se la metaheurística EPSO GIST al dominio multiobjetivo (FEPSO GIST), específicamente aplicada al problema de balance de cargas en un SDEE BT, se presentará un diagrama de flujo que describe su algoritmo soporte.

## 2.2. La metaheurística AFS

Como se dijo en la primera parte del presente trabajo, este algoritmo metaheurístico, referido en el estado del arte como AFSA (*Artificial Fish Shool Algorithm* o Algoritmo de Cardumen de Peces Artificiales), fue propuesto por Xiaolei Li. Se presentaron, en la primera parte, *los cinco comportamientos tipificados del cardumen*. Resta desarrollar el algoritmo. En este trabajo, como aporte, se ha re-formulado la estrategia del algoritmo AFSA, conforme el problema de balance de cargas. Los siguientes desarrollos se apartan algo de su formulación primigenia, a los efectos de brindar mayor compatibilidad con los cinco principios de la inteligencia de grupo.

Definición de los Parámetros y Variables en el AFSA: Sea  $[AF] = \{af_1, af_2, \dots, af_{NF}\}$  el conjunto de peces artificiales, referido como cardumen o población; cada  $af_i$ , ocupará, en el espacio de búsqueda y para cierta instancia de decisión (iteración del algoritmo), una posición  $\mathbf{x}(af_i)$ , la cual resulta un vector cuyas componentes son cada una de las variables de decisión del problema. La *aptitud* que exhibe el pez artificial  $af_i$ , en términos de localización de máximo alimento, ocupando la posición  $\mathbf{x}(af_i)$ , será indicada como  $f(\mathbf{x}(af_i))$ , siendo  $f$  la *función de aptitud* establecida para la estrategia de optimización. Simplificadamente, puede emplearse la notación:  $\mathbf{x}_i \equiv \mathbf{x}(af_i)$  y  $f(\mathbf{x}_i) \equiv f(\mathbf{x}(af_i))$ , porque todo pez artificial siempre ocupa una posición y siempre tiene una aptitud, tal y como se hace en el PSO. Sean, entonces, dos peces artificiales  $\mathbf{x}_i$  y  $\mathbf{x}_j$ , la distancia entre ambos resultará:  $\mathbf{d}_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$ . Además, *visual* es el campo visual o alcance de un  $af_i$  y *step*, la distancia que puede moverse un  $af_i$  en cada paso de iteración;  $\delta$  es el grado de congestión e indica, en el comportamiento de un  $af_i$ , si existe alimento en la vecindad de cierto punto o concentrado en él. Su valor pertenece a  $[0, 1]$ . *Visual*, *step* y  $\delta$  son parámetros. Además se tienen los siguientes: *NF*: Número de peces artificiales o tamaño de la población/cardumen, *TryNum*: es el máximo número de intentos de un pez artificial para encontrar una mejor posición,  $\mathbf{x}_{Best}$ , y será aplicado en el comportamiento prey o de ataque. Adicionalmente, *MaxItera*: es el número máximo de iteraciones del algoritmo, condición estricta de salida del mismo, y *NumEstable* es el

número fijado como límite de iteraciones en los que el registro no es actualizado, por no existir cambios en las aptitudes (y, consecuentemente, posiciones de los  $af_i$  en el cardumen), y supone una condición no estricta de salida del algoritmo. Los comportamientos de ataque, grupal o de seguimiento, se definen en el modelo, y en términos matemáticos, como sigue:

- 1) De ataque: Si, en cierta instancia de decisión,  $\mathbf{x}_i$  es el estado de  $af_i$  y  $\mathbf{x}_j$  es el estado de  $af_j$ , estando  $af_j$  en el campo de visión de  $af_i$ , esto es:  $|\mathbf{d}_{ij}| \leq \text{visual}$ , entonces se procede a comparar las aptitudes  $f(\mathbf{x}_i)$  y  $f(\mathbf{x}_j)$ . Si  $f(\mathbf{x}_i) > f(\mathbf{x}_j)$ ; entonces  $af_i$  se moverá  $\text{Random}(\text{step})$  (función aleatoria de *step*) hacia y en la dirección de  $\mathbf{a}_j$ , dentro de su campo de visión. En caso contrario, intentará un número *TryNum* de veces encontrar una mejor posición,  $\mathbf{x}_{Best}$ , que la que ocupa, dentro de su campo de visión, y, si falla,  $af_i$  se moverá aleatoriamente  $\text{Random}(\text{step})$  dentro de su campo de visión.

Matemáticamente y en pseudo-código:

```

Si  $f(\mathbf{x}_j) > f(\mathbf{x}_i)$  entonces
 $\mathbf{x}_{i,next} = \mathbf{x}_i + \text{Random}(\text{Step}) \times [(\mathbf{x}_j - \mathbf{x}_i) / |\mathbf{x}_j - \mathbf{x}_i|]$ 
Sino
Para  $k=1$  hasta TryNum
 $\mathbf{x}_{i,next} = \mathbf{x}_i + \text{Random}(\text{Step})$ 
Si  $f(\mathbf{x}_{i,next}) > f(\mathbf{x}_i)$  entonces
Fin Para
FinSi
FinPara
 $\mathbf{x}_i,next = \mathbf{x}_i + \text{Random}(\text{Step})$ 
FinSi

```

Siendo  $\text{Random}(\text{Step})$  una función que genera números aleatorios en el intervalo  $[0, \text{step}]$  con distribución uniforme de probabilidad.

- 2) De grupo: Si, en cierta instancia de decisión,  $\mathbf{x}_i$  es el estado/posición de  $af_i$ , *nf* es el número de  $af_i$  en el campo visual de  $af_i$ , dado por la función *N* definida mediante  $nf = N(\mathbf{x}_i, \text{visual}) = \{ \mathbf{x}_j / |\mathbf{d}_{ij}| \leq \text{visual} \}$ , y  $\mathbf{x}_c$  es la posición centro de la vecindad de  $\mathbf{a}_i$  y definida por  $\mathbf{x}_c = \sum_{j=1}^{nf} (\mathbf{x}_j / nf)$ , se tiene, en pseudo-código:

```

Si  $f(\mathbf{x}_c) > f(\mathbf{x}_i)$  y  $(nf/NF) < \delta$  entonces
 $\mathbf{x}_{i,next} = \mathbf{x}_i + \text{Random}(\text{Step}) \times [(\mathbf{x}_c - \mathbf{x}_i) / |\mathbf{x}_c - \mathbf{x}_i|]$ 
Sino
Comportamiento de ataque
FinSi

```

3) De seguimiento: Si, en cierta instancia de decisión,  $\mathbf{x}_i$

es el estado/posición de  $a_i$ , y  $a_{iMax}$  es el af en el campo visual de  $a_i$  que ha alcanzado la máxima *consistencia alimenticia* de la vecindad  $nf$ ,  $\mathbf{x}_{Max}^{nf}$ , siendo  $f(\mathbf{x}_{Max}^{nf})$  su aptitud, se tiene en pseudo-código:

Si  $f(\mathbf{x}_{Max}^{nf}) > f(\mathbf{x}_i)$  y  $(nf/NF) < \delta$  entonces  
 $\mathbf{x}_{i\ next} = \mathbf{x}_i \times \text{Random}(\text{Step}) \times [(\mathbf{x}_i - \mathbf{x}_{Max}^{nf}) / |\mathbf{x}_i - \mathbf{x}_{Max}^{nf}|]$   
 Sino  
 Comportamiento de ataque  
 FinSi

De modo que la pertinencia referida para los comportamientos de grupo y de seguimiento está dada por la misma condición de congestión en la vecindad  $nf$ :  $(nf/NF) < \delta$ . Si esta condición no se satisface, no son pertinentes los comportamientos de grupo o de seguimiento, y tiene lugar el de ataque.

Con estos desarrollos, puede ser definido el Algoritmo AFS, en forma de pseudo-código y por pasos: **Paso 1)** Inicialización de los parámetros NF, visual, TryNum,  $\delta$ , step, MaxItera y NumEstable; **Paso 2)** Inicialización de la población o cardumen: al igual que en la metaheurística PSO o sus variantes, como la EPSO GIST considerada en este trabajo, la población/cardumen debe ser inicializada con valores pertinentes. Es un punto crítico en la aplicación del balance de cargas debido a las singulares características que exhibe el espacio de búsqueda; **Paso 3)** Inicialización del Registro (Bulletin) para evaluar los nuevos estados de cada af y, si procede, actualizarlos, durante el procedimiento iterativo, con los mejores estados alcanzados. En el comienzo, el registro contiene posiciones/estados y aptitudes coincidentes con la inicialización de la población/cardumen; **Paso 4)** Se aplica el comportamiento normativo del cardumen, que consiste en la evaluación del movimiento; esto supone, evaluar la pertinencia de los comportamientos de grupo y de seguimiento, observando si se satisface la condición  $(nf/NF) < \delta$ . De ser pertinentes, proceden ambas, y la posición y aptitud para un nuevo estado proviene del comportamiento que arroje la mayor aptitud.

Si no son pertinentes estos comportamientos, el cardumen adopta el comportamiento de ataque. Es claro que las evaluaciones y cálculos correspondientes, se ejecutan para cada af del Cardumen; **Paso 5)** Actualización del Registro (Bulletin Update): Si procede, se registran los nuevos valores de posición y aptitud para cada af del cardumen, y se actualiza el

número de iteración del proceso. Se evalúa si tiene lugar la condición de parada, consistente en la permanencia del registro sin modificaciones durante el número NumEstable de iteraciones, o bien en alcanzar el número MaxItera de iteraciones. Si una de estas condiciones tiene lugar, se sigue con el paso 6. De lo contrario, se regresa al paso 4, incrementando, previamente en 1, el contador de iteraciones; **Paso 6)** Condición de salida satisfecha: se almacenan posiciones y aptitudes de cada af del cardumen, así como otros parámetros de interés registrados. Desde allí se obtiene el  $\mathbf{x}_{Max}$  y su  $f(\mathbf{x}_{Max})$ .

Se está ahora en condiciones de establecer el interés por esta metaheurística, AFS: analizando los pseudocódigos propuestos y aplicándolos para optimizar los objetivos del problema de balance de cargas individualmente, puede observarse que el movimiento del cardumen propende a la estabilidad, dentro de los principios de la IG.

De modo que ambas metaheurísticas, EPSO GIST y AFS, exhiben características complementarias. De aquí la propuesta de integrarlas en un algoritmo híbrido, matheurística con inicialización MIP, como mejor estrategia de solución para el problema abordado.

### 3. EXTENSIÓN DE LAS METAHEURÍSTICAS EPSO GIST Y AFS AL DOMINIO MULTI OBJETIVO

#### 3.1. Conjuntos difusos

Resulta pertinente abordar el concepto de conjunto difuso, ya que constituye la herramienta central que se adoptó para la extensión de las metaheurísticas propuestas al dominio multiobjetivo.

Un conjunto clásico o rígido exhibe una función binaria de pertenencia de sus elementos. Esto significa que la misma asume, por caso, el valor 1, si un elemento pertenece al conjunto, o el valor 0, si no pertenece.

Por el contrario, L. Zadeh propone, en 1970 [2], la idea de una *función de pertenencia continua*, solidaria a un conjunto, asumiendo valores en el intervalo [0, 1], es decir que, para tal conjunto, pueden establecerse grados de pertenencia, *cuyos valores se encuentren en dicho intervalo*. A este tipo de conjun-

tos los denominó difusos o borrosos (*Fuzzy Sets*). En la Fig. 1, se representa un conjunto difuso lineal, tipo que será empleado para modelar los objetivos (y restricciones, en general), a los efectos de transformar la formulación clásica de un problema de optimización, en su formulación difusa.

En la Fig. 1,  $v$  es una variable genérica (por ejemplo, las pérdidas de potencia activa) y  $\mu(v)$  la *función de pertenencia definida*. El valor  $v_{Min}$  tiene el máximo grado de aceptación,  $\mu(v) = 1$ , al igual que valores inferiores a  $v_{Min}$ . Si  $v > v_{Min}$ , el grado de aceptación comienza a disminuir hasta arribar a  $v = v_{Max}$ , valor que resulta inaceptable,  $\mu(v) = 0$ , así como los valores  $v > v_{Max}$ . Se observa que puede ser definido también un nivel o grado de aceptación,  $\alpha = \alpha_c$ , en cuyo caso serán aceptables valores en el intervalo  $[v_{Min}, v(\alpha_c)]$ . A tal intervalo, se lo refiere como  $\alpha_c$ -corte del CD.

### 3.2. Toma de decisión estática difusa

Para la *toma de decisión estática difusa*, Bellman y Zadeh [2], introducen el concepto de conjunto difuso de decisión. Si se consideran  $n$  conjuntos difusos, queda definido mediante la expresión:

$$\{D\} = \{C_1\} <opC> \{C_2\} <opC> \dots <opC> \{C_{n-1}\} <opC> \{C_n\} \quad (3)$$

donde  $<opC>$  es un operador entre conjuntos difusos que recibe el nombre de *confluencia*. La *confluencia* más frecuentemente empleada es la *intersección*. Asociado al operador  $<opC>$  entre los conjuntos difusos, existe un operador matemático,  $opC$ , aplicable a sus *funciones de pertenencia*, que genera, desde (3), el valor de pertenencia resultante en el conjunto difuso de decisión,  $\{D\}$ , es decir:

$$\mu\{D\} = \mu\{C_1\} opC \dots opC \mu\{C_{n-1}\} opC \mu\{C_n\} \quad (4)$$

Al trabajar con la confluencia  $<opC> \equiv \cap$ , el operador  $opC$  recibe el nombre general de  $t$ -norma. Una  $t$ -norma es, entonces y para este contexto, una función  $t$ , de las *funciones de pertenencia* solidarias a cada conjunto difuso  $\{C\}$ ,  $t = f\{\mu\{C_1\}, \mu\{C_1\}, \dots, \mu\{C_n\}$ , definida formalmente como sigue:

Una  $t$ -norma es una *función*  $t$ , definida en el intervalo  $[0, 1]$  y aplicada también en  $[0, 1]$ , que satisface las siguientes condiciones: si  $t: [0, 1] \rightarrow [0, 1]$ , entonces: **a.-**  $t(0,0) = 0$ ;  $t(x,1) = x \rightarrow$  *Condiciones de frontera*;

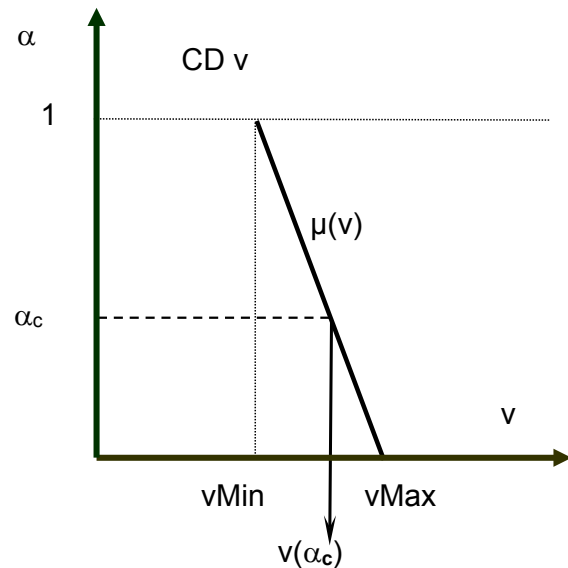


Fig. 1. Conjunto difuso triangular (CDT) definido sobre la variable  $v$

**b.-**  $t(x,y) = t(y,x) \rightarrow$  *Conmutatividad*; **c.-** si  $x \leq \alpha$  e  $y \leq \beta$   $t(x,y) \leq t(\alpha,\beta) \rightarrow$  *Monotonicidad* y **d.-**  $t(t(x,y),z) = t(x,t(y,z)) \rightarrow$  *Asociatividad*.

Entonces, establecida la  $t$ -norma a emplear,  $t$ , la decisión maximizante estática en el conjunto difuso de decisión, para cierta instancia, resultado de la *confluencia* – intersección – entre los conjuntos difusos solidarios a cada objetivo/restricción del problema, vendrá dada por la expresión:

$$\mu\{D\}_{Max} = \text{MAX} \{ \mu\{C_1\} t \dots t \mu\{C_{n-1}\} t \mu\{C_n\} \} \quad (5)$$

De modo que, si se tiene un problema de optimización clásico multiobjetivo y restringido, con  $n$  funciones objetivo (FO) y  $m$  restricciones (R), su solución en el dominio difuso requerirá de: a) Definir un conjunto difuso para cada FO en  $[1..n]$  y R en  $[1..m]$  y b) Determinar la decisión maximizante, según (5).

Cada instancia de decisión -iteración algorítmica- arrojará un valor de  $\mu\{D\}$ . En tal proceso, se identifica el  $\mu\{D\}_{Max}$ . Cuando por un número definido de iteraciones  $\mu\{D\}_{Max}$  no se modifique, se habrá arribado a la *solución más satisfactoria*. Y desde los valores  $\mu\{C_1\}, \mu\{C_2\} \dots \mu\{C_n\}$ , se recupera, dado que cada  $\{C_i\}$  está asociado a una variable objetivo/restricción, el *vector solución más satisfactorio en las variables de control y de estado*.

Para el problema del balance de cargas en un SDEE BT, se tienen como objetivos: 1) Minimizar las pérdidas de potencia activa (kW); 2) Minimizar el valor de

la componente homopolar (A), 3) Maximizar el índice de tensión de suministro y d) Minimizar el número cambios de fase. El 4) es la variable de control, mientras que 1), 2) y 3) resultan variables de estado.

Cada uno de los conjuntos difusos, que serán definidos en sección siguiente, puede tener una importancia que se corresponda con la prioridad establecida por el tomador de decisiones para los objetivos.

Para ello se introducen los denominados ponderadores exponenciales o de Yager [4],  $pY$ , que son números reales positivos, tal que su suma sea igual al número de conjuntos difusos involucrados en la toma de decisión. A mayor valor de  $pY_i$ , mayor importancia del objetivo  $FO_i$ , cuyo conjunto difuso asociado es  $\{C_i\}$ . Operacionalmente, la *función de pertenencia*  $\mu\{C_i\}$  se eleva al valor  $pY_i$ , de modo que se tiene una *función de pertenencia ponderada*,  $\mu\{C_i\} pY_i$ . Y (5) se transforma en:

$$\mu\{D\}_{Max}^{pY} = \text{MAX} \{ \mu\{C_1\} pY^1 \text{ t...t } \mu\{C_{n-1}\} pY^{n-1} \text{ t } \mu\{C_n\} pY^n \} \quad (6)$$

El efecto de  $pY_i$ , resulta en una *contracción* de  $\{C_i\}$ , si  $pY_i > 1$  (mayor importancia en la *confluencia*), o una *dilatación* de  $\{C_i\}$ , si  $pY_i < 1$  (menor importancia en la *confluencia*). Se muestra en la Fig. 2.

#### 4. MODELO DIFUSO PARA EL PROBLEMA

Se emplea, para todos los objetivos, el conjunto difuso representado en la Fig. 1, el cual se define, para la variable genérica  $v$ , del siguiente modo:

$$\mu(v) = 1; \text{ si } v < v_{Min} \quad (7)$$

$$\mu(v) = 0; \text{ si } v > v_{Max} \quad (8)$$

y si  $v_{Min} \leq v \leq v_{Max}$

$$\mu(v) = [(v_{Max}-v)/(v_{Max}-v_{Min})]^{pY} \quad (9)$$

En el caso de las pérdidas de potencia activa y la componente homopolar, se trata de magnitudes obtenidas desde herramientas del análisis de funcionamiento de los sistemas de potencia (fundamentalmente, el flujo de potencia trifásico). En el caso del número de cambios de fase, es la variable de control, de modo que la estrategia de solución será quien lo contabilice. Pero en el caso de la tensión de suministro, se trata de un índice que debe ser definido.

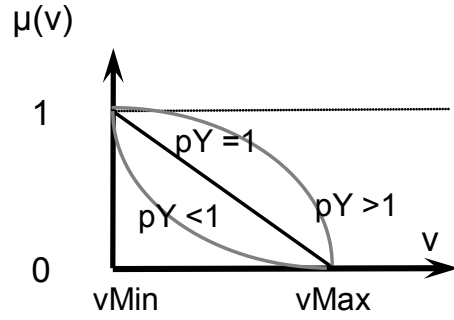


Fig. 2. Efecto del ponderador de Yager sobre un conjunto difuso

Se propone la siguiente formulación:

a) Se fijan: un límite inaceptable de tensión en los nodos terminales,  $u_{InAd}$ , y un límite aceptable en los mismos,  $u_{Ad}$ . Los empleados en la simulación son:  $u_{lnad}=0.92$  [pu] y  $u_{ad} = 0.95$  [pu]; b) Se definen los valores recíprocos de las tensiones límites especificadas en a):  $v1=1/u_{Ad}$ , y  $v2=1/ u_{lnAd}$ ; c) Dado un valor de tensión,  $u$ , en cierto nodo terminal,  $nt$ , se calcula su recíproco,  $vft=1/u$ . Luego, la expresión condicional para la función de pertenencia correspondiente al conjunto difuso, equivalente a un valor de tensión fuera de tolerancia variable  $vft$ , evaluada en el nodo terminal  $nt$ , resulta:

$$\mu_{nt}(vft) = 1; \text{ si } vft < v1 \quad (10)$$

$$\mu_{nt}(vft) = 0; \text{ si } vft > v2 \quad (11)$$

y si  $v2 \geq vft \geq v1$

$$\mu(vft) = [(v2-vft)/(v2-v1)]^{pYvft} \quad (12)$$

Finalmente, como índice de tensión fuera de tolerancia, se propone la media geométrica de todos los valores  $\mu_{nt}(vft)$  en el conjunto de los *nodos terminales*,  $\{NT\}$ , cuyo número es  $nNT$ :

$$\mu(U_{tf}) = \sqrt[nNT]{\prod_{nt=1}^{nNT} \mu_{nt}(vft)} \quad (13)$$

Como se dijo, se trata de un índice estricto, puesto que no se admite que tan solo 1 nodo terminal tenga una tensión de suministro inadmisibles. Para la operación entre las funciones de pertenencia ponderadas, ecuación (6), correspondientes a los conjuntos difusos, se propone la t-norma llamada Producto de Einstein y definida como, siendo  $x$  e  $y$  dos *funciones de pertenencia* genéricas:

$$t_{PEinstein}(x,y) = \frac{xy}{1 + (1-x) + (1-y)} \quad (14)$$

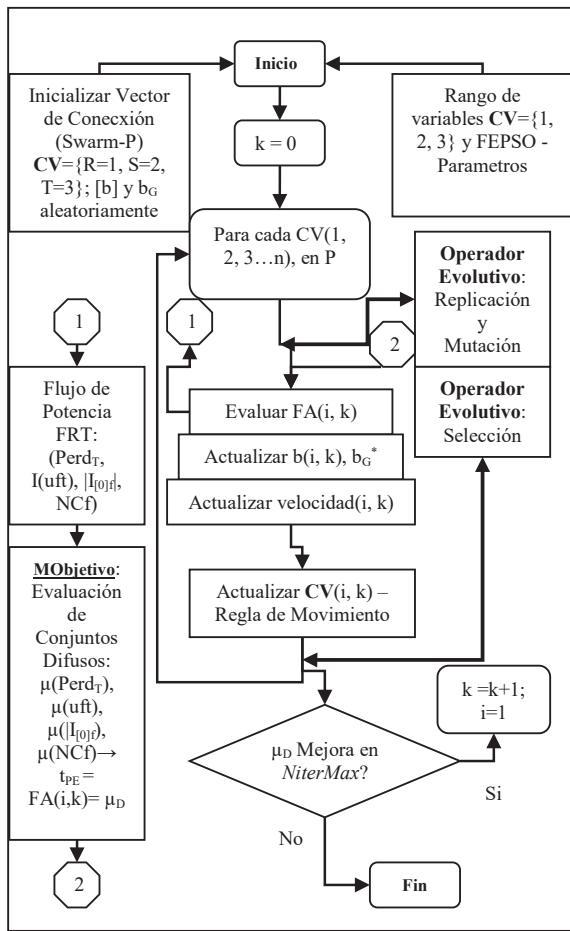


Fig. 3. Diagrama de flujo para la metaheurística FEPSO GIST

La  $t_{PEinstein}$  se aplica tomando un par de funciones y, luego, su resultado es una función que se opera con la que sigue, y así continúa el procedimiento. En la Fig. 3, se muestra la estrategia de solución, empleando solo la metaheurística FEPSO GIST.

La aptitud de cada partícula FEPSO GIST, o pez artificial FAFS, en el dominio multiobjetivo, viene dada, en cada iteración  $k$ , por la función  $t$ -norma  $t_{PEinstein}$ . Al extender las metaheurísticas propuestas al dominio multiobjetivo, sugen tres requerimientos adicionales a la satisfacción de los principios IG: a) Captación de incertidumbres de valor en las variables, aspecto resuelto mediante los propios conjuntos difusos, pues les es una cualidad inherente; b) Compatibilidad métrica, aspecto que refiere la capacidad de que la estrategia de solución pueda localizar buenas soluciones en el espacio homónimo, conforme sus características, en la *confluencia* de los objetivos, resuelto por la topología FEPSO GIST, y la integración en una

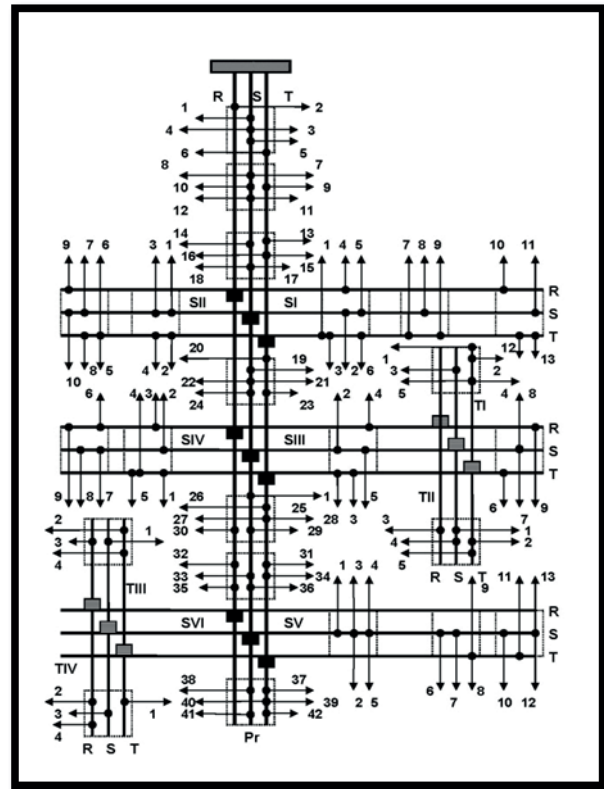


Fig. 4. Esquema Trifilar del SDEE BT real analizado

estrategia matheurística (Math) de solución FEPSO GIST, FAFS y programación lineal entera mixta, MIP, como técnica de *arranque o bootstrapping*; y c) La Pareto-dominancia en las mejores soluciones obtenidas, aspecto sustancial en la optimización multiobjetivo resuelto por el empleo de la  $t$ -norma Producto de Einstein, como función de aptitud, FA.

### 5. SIMULACIÓN Y RESULTADOS

El esquema del SDEE BT real empleado para la simulación, descrito en la primera parte del presente trabajo, se presenta en la Fig. 4.

Los valores de las variables de estado previas a la optimización, resultaron:  $\{ Perd_T = 13.02 \text{ (Kw)}; |I_{0j}^{[0]}| = 47.6 \text{ (A)}; I(\Delta u) = 0 \}$ , siendo  $I(\Delta u)$ , el índice de tensión de suministro. Las variables de estado y control (NC = Nro. de Cambios de Fase) resultaron, para una población  $P = 200$  (partículas o peces artificiales): a) FEPSO GIST =  $\{ Perd_T = 7.21 \text{ (kW)} \}; |I_{0j}^{[0]}| = 0.4 \text{ (A)}; I(\Delta u)_{Ref} = 0.27; NC = 59$ ; y b) Math =  $\{ Perd_T = 6.12 \text{ (kW)} \}; |I_{0j}^{[0]}| = 0.2 \text{ (A)}; I(\Delta u)_{Ref} = 0.54; NC = 57 \}$ . Los resultados logrados mediante la me-

taheurística FSTO con una técnica de bootstrapping basada en el algoritmo BI, fueron:  $FSTO+BI = \{ Perd_T = 9.21 \text{ (kW)}; |I_{S,Ref}^{[0]}| = 5.8 \text{ (A)}; I(\Delta u)_{Ref} = 0.1\}$ ;  $NC = 73$  }, claramente inferiores en calidad. Los valores  $vMin$  y  $vMax$  para la construcción de los conjuntos difusos solidarios a las variables  $Perd_T$ ,  $|I_{S,Ref}^{[0]}|$  y  $NC$ , se obtienen:  $vMin$  desde optimizaciones individuales FSTO GIST; y  $vMax$ , desde un porcentaje de  $vMin$ , sumándosele.

## 6. TRABAJOS FUTUROS

Las líneas de investigación futuras sobre estas estrategias de optimización son múltiples. La seguida por el autor del trabajo, se refiere a las hiperheurísticas, estrategias de más alto nivel, capaces de comandar la selección de meta/matheurísticas más aptas, conforme evoluciona la búsqueda de soluciones.

## 7. CONCLUSIONES

A. Se ha presentado una estrategia de solución metaheurística, primero, e híbrida o matheurística, luego, mediante novedosos aportes, orientadas a resolver un complejo problema combinatorio: El balance de cargas en un SDEE BT. La topología GIST aplicada sobre la metaheurística FSTO, el empleo complementario de la AFS, extendidas al dominio multiobjetivo mediante el empleo de conjuntos difusos, y la concepción de un modelo matheurístico como el descrito, arroja excelentes resultados.

B. La estrategia matheurística consiste en emplear el modelo MIP, presentado en la primera parte del trabajo, obtener un conjunto inicial de soluciones factibles, para luego aplicar, modificando la característica de la carga desde intensidad constante a potencia constante, la metaheurística FSTO GIST hasta detectar pocos cambios en la *función de aptitud*, y pasar a la aplicación de la metaheurística FAFS.

## REFERENCIAS

[1] G.A. Schweickardt, "Metaheurísticas Multiobjetivo Cardumen de Peces Artificiales (FAFS) y Optimización Evolucionaria por Enjambre de Partículas

con Topología Estocástica Global Individual (FEPSTO GIST). Parte I: Antecedentes y Desarrollos Teóricos", *Lámpsakos*, N° 12, pp. 13-22, 2014.

[2] G. Schweickardt & V. Miranda, "Metaheurística FSTO aplicada a problemas de optimización combinatoria: Balance de fases en sistemas de distribución eléctrica". *Revista Ciencia, Docencia y Tecnología. Universidad Nacional de Entre Ríos*. Nro. 40, pp. 133-16, 2010.

[3] R. Bellman & L. Zadeh, "Decision-making in a fuzzy environment". *Management Science*. Nro. 17, pp. 141-164, 1970.

[4] R. Yager, "Multiple objective decision making using fuzzy sets". *Intl. J. Man-Machine Studies*. 9, pp. 53-64, 1977.

[5] A. Díaz, F. Glover, H. Ghaziri & J. González, "Optimización heurística y redes neuronales", Paraninfo, Madrid, España, 1996.

[6] F. Glover. "Tabu Search: Part I". *ORSA Journal on Computing*. Vol. 1, pp. 190-206, 1989.

[7] P. Black, "Dictionary of algorithms and data structures". National Institute of Standards and Technology, 2009. Disponible: <http://www.nist.gov/index.html>.

[8] I. Osman & J. Kelly, "Meta-Heuristics: Theory and Applications". Ed. Kluwer Academic. Boston, USA, 1996, .

[9] G. Beni & J. Wang, "Swarm intelligence in cellular robotic systems", *Proceedings NATO Advanced Workshop on Robots and Biological Systems*, Tuscany, Italy, June, 1989.

[10] J. Kennedy & R. Eberhart, "Particle swarm optimization". *IEEE International Conference on Neural Networks*, IEEE Service Center, Piscataway, NJ. Pert, Australia, 1995.

[11] X. Li, Z. Shao & J. Qian, "An optimizing method based on autonomous animats: Fish-swarm algorithm", in *Systems Engineering Theory and Practice*. Vol. 22, no. 11, pp. 32-38, 2002.



# Análisis de los Indicadores de Emisiones de Gases de Efecto Invernadero para Valorar Proyectos de Energía Renovable en Sistemas Eléctricos, Caso de Colombia

Analysis Indicators of Greenhouse Gas Emissions for Assessing Renewable Energy Projects in Electrical Systems, Case of Study in Colombia

**Marco Alejandro Berrío-Monsalve, MSc**

*Facultad de Minas  
Universidad Nacional de Colombia  
Medellín, Colombia  
maberrio@unal.edu.co*

**Mónica Arango-Arango, MSc**

*Facultad de Administración  
Universidad de Medellín  
Medellín, Colombia  
moarango@udem.edu.co*

**Sergio Botero-Botero, PhD**

*Facultad de Minas  
Universidad Nacional de Colombia  
Medellín, Colombia  
sbotero@unal.edu.co*

(Recibido el 30-11-2014. Aprobado el 13-12-2014)

**Resumen.** La contribución técnica presenta una discusión acerca los indicadores definidos por la UNFCCC y apropiados para aplicar en proyectos de generación eléctrica en Latinoamérica. Se hace énfasis en el posible uso de indicadores para incentivar el uso de tecnologías de generación renovables y mitigar la emisión de gases con efecto invernadero a la atmósfera, en proyectos que se encuentren en fase operativa. Adicionalmente, se presenta un nuevo indicador (El del Costo Social del Carbono- SCC-) diferente a los definidos por la UNFCCC. Finalmente se centra la discusión en la aplicación de los indicadores en el caso Colombiano y cómo la correcta valoración de las emisiones de CO<sub>2</sub> permite una mayor penetración a las tecnologías de generación renovables como la eólica.

**Palabras clave:** Indicadores de emisiones, valoración, mitigación de emisiones de CO<sub>2</sub>, proyectos de energías renovables.

**Abstract.** The paper presents a discussion about the indicators defined by the UNFCCC and suitable for application in power generation projects in Latin America. Emphasis on the possible use of indicators is done to encourage the use of renewable generation technologies and reduce the emission of greenhouse gases into the atmosphere, in projects that are in operational phase. Additionally, a new indicator (The Social Cost of Carbon-SCC) different from those defined by the UNFCCC is presented. Finally, the discussion focuses on the application of indicators in the Colombian case and how the correct assessment of CO<sub>2</sub> emissions allow greater penetration of renewable technologies such as wind generation.

**Keywords:** Emission indicators, assessment, mitigation of CO<sub>2</sub> emissions, renewable energy projects.

**Citación de artículo, estilo IEEE:**

M.A. Berrío-Monsalve, M. Arango-Arango, S. Botero-Botero, "Análisis de los Indicadores de Emisiones de Gases de Efecto Invernadero para Valorar Proyectos de Energía Renovable en Sistemas Eléctricos, caso de Colombia", *Lámpsakos*, N° 13, pp. 24-38, 2015.

DOI: <http://dx.doi.org/10.21501/21454086.1496>



## 1. INTRODUCCIÓN

Este artículo se deriva del trabajo final de maestría en Ingeniería Administrativa “*Diversificación de la Canasta Eléctrica Colombiana Considerando Costos de Emisiones de CO<sub>2</sub>*”, de la Facultad de Minas, Universidad Nacional de Colombia- Sede Medellín [1]. En los últimos años, se ha visto la necesidad de introducir más proyectos de energías renovables en los sistemas eléctricos, es decir, proyectos diferentes a los de energías convencionales (grandes hidroeléctricas y termoeléctricas con combustibles fósiles) [2], los cuales ayuden a diversificar la matriz de generación eléctrica. Esto es debido a las implicaciones que ha tenido en el cambio climático la emisión de gases de efecto invernadero, los cuales son causados, en gran parte, por el consumo de combustibles fósiles en las centrales termoeléctricas. Sin embargo, a la hora de evaluar proyectos se han encontrado diferentes fuentes de información que pueden llegar a ser confusas [3].

Resulta claro que los sistemas eléctricos deben buscar reducir las emisiones de gases de efecto invernadero. Esto se logra sustituyendo las tecnologías que consumen combustibles fósiles por energías renovables; de esta forma, se contribuiría significativamente a reducir los impactos del calentamiento global.

Los sistemas eléctricos alrededor del mundo, tanto los centralizados como los que tienen una estructura de mercado, despachan su energía en orden de mérito, dando prelación a las de menores costos. Sin embargo, en pocos sistemas a las termoeléctricas se les incluye el costo de las emisiones de gases de efecto invernadero que ellas ocasionan. Si se considera el costo por contaminar dentro de la decisión de despacho, se tendría como resultado que las energías no convencionales mejorarían su competitividad con respecto a las termoeléctricas y las desplazarían, logrando una reducción de emisiones. La Organización para la Cooperación y el Desarrollo Económico (OCDE) le presentó al gobierno colombiano en 2015 una serie de recomendaciones sobre aspectos económicos. En el informe, la OCDE hace énfasis en introducir un impuesto directo a las emisiones de CO<sub>2</sub> [4], el cual sería incluido en una futura reforma tributaria.

En este artículo se analizan los diferentes índices derivados a partir de los mecanismos de desarrollo limpio –MDL en español y CDM en Inglés- (Clean Development Mechanism –CDM-). Dichos índices son

utilizados para incentivar los nuevos desarrollos en generación renovable y/o para valorar las emisiones de CO<sub>2</sub> en proyectos existentes, lo cual, a su vez, incentiva sus reducciones.[5], [6].

Se hace especial énfasis en los CER (Certificados de Reducción de emisiones) y los EUA (European Union Allowances) [6], además de los trabajos de valoración del costo social del carbono (Social Cost of Carbon) [7], [8]. Finalmente se identifican los tipos de proyectos para los cuales deberían ser aplicados.

Hablar de CDM necesariamente implica hacer un recuento desde el Protocolo de Kyoto sobre las medidas tomadas a nivel global y los mecanismos desarrollados para mitigar los efectos del cambio climático. El protocolo fue inicialmente adoptado por 186 países, el 11 de diciembre de 1997, en Kyoto, Japón, y entró vigencia a partir del 16 de febrero de 2005. Es importante mencionar que la implementación del protocolo se planeó en tres fases así:

### 1.1. Fases de implementación del protocolo de Kyoto

Los expertos que participaron en la elaboración del protocolo de Kyoto [9] visualizaron la dificultad de su implementación a futuro, motivo por el cual la dividieron en varias fases. La ventaja de dividir la implementación en diferentes fases consiste en que se detectan y corrigen los problemas o dificultades que surgieron en las primeras fases. A continuación se describe cada una de las fases propuestas.

#### 1.1.1. Fase I

Esta inició el 16/02/2005 y terminó el 31/12/2007. Se considera una fase de aprendizaje en la cual fueron creados cuerpos administrativos y regulatorios. Durante esta fase, 11000 de las mayores industrias contaminantes, incluidas centrales de generación térmica en la Unión Europea (UE), estaban obligadas a suministrar datos sobre sus niveles de emisiones anuales. Con base en los datos suministrados por las industrias, los gobiernos hicieron una emisión de “allowances” o permisos para emitir toneladas de CO<sub>2</sub>. En el año 2006, se observó que la emisión de “allowances” o permisos fue excesiva, lo cual colapsó los precios de dichos títulos en el mercado por sobreoferta. Dhamija [3] afirma que la sobreoferta de permisos para emitir CO<sub>2</sub> se dio debido a que las grandes industrias y centrales de generación térmica

no fueron totalmente transparentes acerca de los datos suministrados a los organismos gubernamentales sobre los niveles de emisiones “reales”. Adicional al colapso en los precios de los títulos, la consecuencia práctica fue que los grandes emisores de CO<sub>2</sub>, no se vieron obligados a iniciar la reducción de las emisiones durante esta primera fase.

### 1.1.2. Fase II

Esta inició el 01/01/2008 y terminó el 31/12/2012. En este período se corrigió el problema de la sobreoferta de títulos correspondientes a las emisiones de CO<sub>2</sub> y sus equivalentes. La corrección consistió en hacer que todos los títulos emitidos en la primera fase expiraran con el fin de la misma. Para la segunda fase, se emitieron nuevos títulos basados en el nivel de emisiones real verificado correspondiente a las industrias en la Unión Europea (UE); de esa manera, no se generó sobreoferta. En la segunda fase las industrias sí se vieron obligadas a reducir sus emisiones anuales o a pagar por contaminar.

### 1.1.3. Fase III

Esta inició el 01/01/2013 y termina el 31/12/2020. Es la fase actual en la cual los mercados de emisiones de carbono están afrontando serias dificultades por razones que serán explicadas posteriormente. Es importante mencionar que los títulos emitidos por los diferentes países europeos más Rusia durante la Fase II y que no hayan sido acreditados o “consumidos” por las industrias contaminantes, no expiran al final de la fase II. Esto quiere decir que dichos títulos emitidos durante la Fase II se pueden transar durante la Fase III hasta que sean acreditados o “consumidos”.

## 1.2. Anexos del Protocolo de Kyoto

Los 191 países -actualmente- firmantes del protocolo [9] están divididos en tres categorías así:

### 1.2.1. Anexo I

Este grupo de países está integrado por 41 países industrializados, los cuales están obligados a cortar sus emisiones de gases con efecto invernadero en 5,2%, de manera conjunta, por debajo de los niveles de emisión de 1990. Esta reducción se debe llevar a cabo durante el desarrollo de la Fase II. Adicionalmente, este grupo de países está obligado a ayudar

a “bloquear” los niveles de emisiones en las naciones en desarrollo, de tal forma que no se incrementen descontroladamente en la medida en que se incrementa la demanda energética, como consecuencia del desarrollo económico y social.

### 1.2.2. Anexo II

Este grupo está conformado por 24 de los 41 países industrializados más ricos, los cuales están obligados a proveer financiamiento y soporte técnico adicional a las naciones No anexas. Las naciones No anexas corresponden a los 145 países en desarrollo, que necesitan ayuda y transferencia tecnológica para poder “bloquear” sus emisiones de CO<sub>2</sub>.

### 1.2.3. Países No anexados a I

Estas naciones no tienen metas específicas a cumplir respecto de sus niveles de emisiones de CO<sub>2</sub>. Esto se debe a que los países en desarrollo tienen una contribución de emisiones per cápita de CO<sub>2</sub> mínima al compararse con las naciones industrializadas o desarrolladas.

La “United Nations Framework Convention on Climate Change” (UNFCCC) adoptó el principio de responsabilidades comunes para disminuir los niveles de emisiones de gases con efecto invernadero, pero diferenciadas según la capacidad que tiene cada país, de acuerdo a su nivel de desarrollo económico y tecnológico [3]. En este principio se basan los mecanismos flexibles para el desarrollo de tecnologías e instrumentos financieros que permiten concretar las metas inicialmente propuestas por el protocolo de Kyoto.

## 1.3. Desarrollo de mecanismos flexibles

El protocolo definió cuatro mecanismos flexibles, los cuales se explican a continuación con base en el glosario de términos [10] definidos por la (UNFCCC).

### 1.3.1. Clean Development Mechanism (CDM) o Mecanismos de Desarrollo Limpio (MDL)

Mecanismos de desarrollo limpio (MDL), por sus siglas en Inglés. Los CDM impulsan y financian los desarrollos tecnológicos con energías renovables, la reforestación y conservación de bosques naturales o cultivos forestales. Un objetivo de los CDM consiste en compensar las emisiones causadas en países in-

dustrializados mediante la captura de CO<sub>2</sub> que realizan los bosques financiados y certificados en Latinoamérica, África y Asia. El segundo objetivo consiste también en compensar las emisiones causadas en países industrializados mediante el financiamiento de desarrollos energéticos con fuentes renovables -es decir que no utilicen combustibles fósiles-. Los Certified Emission Reduction (CER), por sus siglas en inglés, son instrumentos financieros emitidos por Europa, Estados Unidos y Japón, los cuales se utilizan para financiar los proyectos anteriormente mencionados. Las corporaciones que adquieren CER, básicamente están pagando una compensación mediante la financiación de proyectos en el hemisferio sur, por la contaminación que dichas corporaciones causan en el Norte.

### 1.3.2. *International Emission Trading (IET)*

Mercado global para los instrumentos financieros derivados de las emisiones de carbono. Este mercado depende de cantidades de unidades asignadas a cada país y básicamente ha fracasado debido a la posición de los Estados Unidos (empresarios y gobierno).

### 1.3.3. *Joint Implementation (JI)*

Esta produce los instrumentos financieros, Emission Reduction Unit (ERUs), instrumentos que cumplen la misma función que los CER. La diferencia consiste en que son exclusivos para el financiamiento de proyectos de energías renovables, cuidado o cultivo de bosques en Rusia, Ucrania y los antiguos países de la cortina de hierro.

### 1.3.4. *European Union Emission Trading Scheme (EU ETS)*

Este es el mayor mercado de carbono en el mundo y en el cual principalmente se transan los European Union Allowances (EUA), los CER y los ERU. Los EUA son los instrumentos financieros que le dan a su tenedor el derecho de emitir una tonelada de CO<sub>2</sub> o sus gases equivalentes. Cada gran contaminador en Europa, bien sea industria o central térmica de generación, tiene que acreditar (a finales de marzo de cada año) un EUA por cada tonelada de CO<sub>2</sub> que emitió durante su operación el año inmediatamente anterior [11]. En caso tal de que el contaminante esté "largo" entre el número de EUA que posee versus el número de toneladas emitidas, debe pagar fuertes

multas que oscilan entre 100 y 130 euros por cada tonelada de CO<sub>2</sub> emitida y no acreditada con su respectivo EUA.

### 1.3.5. *Clasificación de Títulos CER, agrupados por tipos de tecnologías*

La UNFCCC, en su más reciente publicación para la consulta sobre los CDM y sus metodologías [6], agrupa diferentes caracterizaciones aplicadas por tipos de tecnologías. En la Figura 1 se observan los títulos CER definidos para aplicar únicamente en el desarrollo de nuevos proyectos de expansión en generación con energías renovables. Los títulos ACM002 aplican para la financiación de hidroeléctricas con capacidad mayor de 20MW. Los títulos AMS-I.D. aplican para la financiación de hidroeléctricas con capacidad menor de 20MW. Los títulos AMS-I.B. aplican para la financiación de proyectos de generación eólica. Cabe anotar que los títulos con mayor volumen de transacción son los relacionados con centrales hidroeléctricas. En total, se tienen trece grupos de títulos CER, los cuales se definen de acuerdo al tipo de tecnología a aplicar en proyectos específicos. En la Figura 2, se observan los títulos CER definidos en la categoría de reforestación y otras metodologías, los cuales son utilizados para financiar la reforestación y el cuidado de bosques que garanticen la compensación de emisiones de CO<sub>2</sub> en otras regiones.

## 2. METODOLOGÍA

Una vez hecha la revisión acerca de los CDM creados a partir del protocolo de Kyoto, cobra importancia analizar las dificultades que se presentan para su implementación.

Inicialmente se hace un análisis de los condicionantes históricos de estos índices. Posteriormente se hace un análisis de la serie de tiempo de estos indicadores, identificando las variaciones que han tenido desde su creación y se discute la aplicabilidad de estos [7], para la financiación de proyectos en los sistemas eléctricos de América Latina. Finalmente se presenta una propuesta para valorar el costo de las emisiones de CO<sub>2</sub>, en un estudio de caso del mercado de energía en Colombia.

Este estudio se basa en la valoración de los costos de generación de las plantas termoeléctricas a carbón y gas, en comparación con la energía eólica.

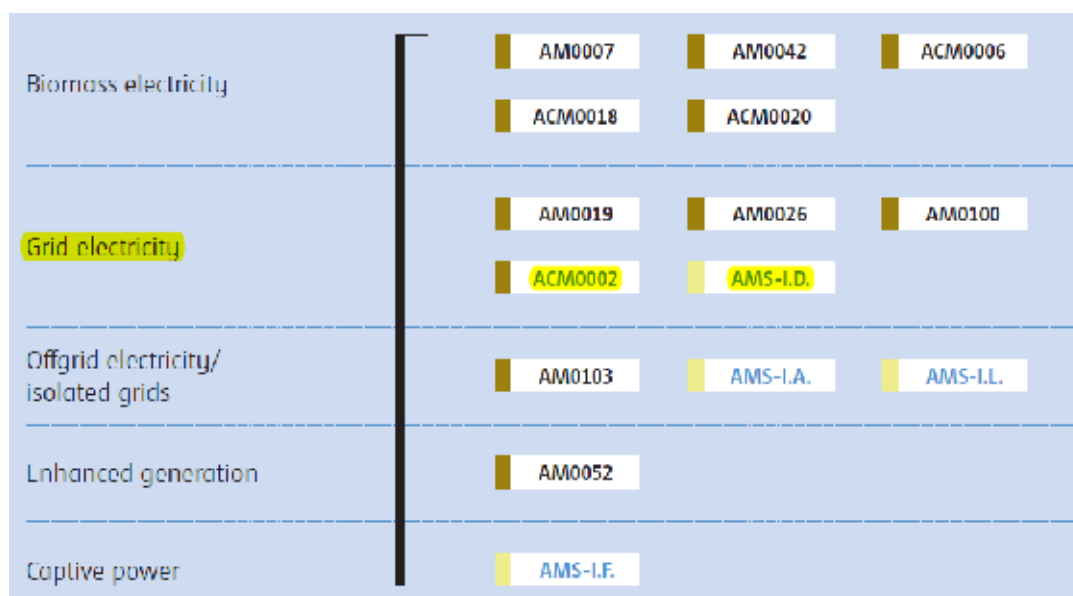


Fig. 1. "Títulos CER" definidos para proyectos con energías renovables, tomado de [6]

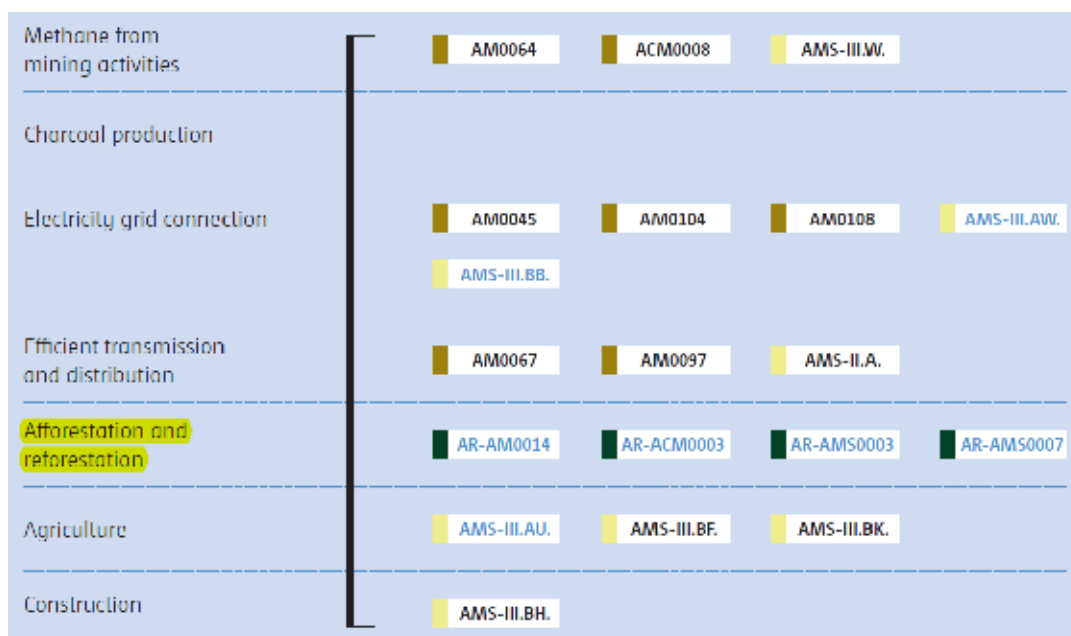


Fig. 2. Títulos CER definidos para reforestación y otras metodologías [6]

## 2.1. Desacuerdos y problemas de implementación del protocolo de Kyoto

Desafortunadamente, China e India -que actualmente se han convertido en grandes contaminantes globales, debido a su incremento en la demanda energética, como consecuencia de su acelerado desarrollo económico-, no tienen limitaciones específicas en sus emisiones de carbono. Este error inicial se debió a que estos dos países tenían un precario desarrollo industrial (por lo cual, no eran grandes contaminadores) durante el período previo a la adopción del protocolo.

Mientras que EU reconoce que los países industrializados son responsables por 70% de la emisión de gases con efecto invernadero, le preocupa simultáneamente el surgimiento de futuras fuentes de dichos gases. De acuerdo a Eizenstat, "Alrededor de 2015, China será el mayor emisor de gases con efecto invernadero y, para 2025, los países en desarrollo en su conjunto emitirán más gases con efecto invernadero, en comparación con el total de los países desarrollados" [12].

El hecho de no haberle exigido a China e India compromisos y metas puntuales en la reducción de emisiones de CO<sub>2</sub>, le ha servido como excusa a Estados Unidos para ser reacio a involucrarse de manera decidida en el mercado de emisiones de carbono. Adicionalmente, el gobierno y el empresariado americano han priorizado el desarrollo económico sobre el problema del cambio climático. La posición del Gobierno americano se ve reflejada en el hecho de que no ha sido lo suficientemente exigente al autoimponer metas ambiciosas en disminución de emisiones de CO<sub>2</sub> a sus industrias más contaminantes e intensivas en consumo energético y a las centrales de generación térmica.

Algunas de las metas definidas por el protocolo son: 8% en reducción de emisiones para la Unión europea como conjunto, 7% para los estados Unidos, 6% para Japón, 0% para Rusia (otro gran error desde el punto de vista ambiental); adicionalmente, permite incrementos de las emisiones de 8% para Australia y 10% para Islandia [9]; esto sin mencionar la contaminación causada por los sistemas de transporte. La nueva sobreoferta de instrumentos financieros CER, ERU y EUA es causada principalmente por la Federación Rusa. El precio de estos instrumentos que actualmente -es decir, durante la Fase III- se transan en el mercado de emisiones europeo ha colapsado, al

punto de que se están transando a precios demasiado bajos, inclusive por debajo de sus costos de emisión. La explicación de la sobreoferta es la siguiente: el colapso económico y político de la Unión Soviética en 1991, ocasionó una gran desindustrialización en Rusia y los países de la antigua cortina de hierro. Esta gran desindustrialización, a su vez, causó una gran caída en la demanda eléctrica y, por ende, una disminución en las emisiones de CO<sub>2</sub> hasta del 40% respecto de los niveles de 1990. Posteriormente, el desarrollo económico ruso ha causado un modesto incremento del 15% de sus emisiones de CO<sub>2</sub> a partir de 1998 [3]. Se puede afirmar entonces que el enorme decremento en las emisiones de CO<sub>2</sub> por parte de Rusia se debe a un efecto colateral, causado por su colapso económico de 1991, más que a un esfuerzo por reducir sus emisiones mediante la implementación de tecnologías de generación limpia o renovable. Como ya se había mencionado, la meta adquirida por los rusos en reducción de emisiones de acuerdo al Protocolo de Kyoto es del 0% y han tenido una reducción real superior al 20%, sin "esforzarse realmente" para alcanzar dicha meta.

La reducción real de emisiones por parte de la Federación Rusa le permitió acumular un estimado de hasta 5 billones de títulos CER, ERU y EUA, -créditos para contaminar- no utilizados durante la Fase II, los cuales son trasladables y transables durante la Fase III. Esta cantidad es el equivalente al esfuerzo hecho por toda la Unión Europea para reducir las emisiones de CO<sub>2</sub> durante la Fase III. Debido a que el Protocolo de Kyoto es vinculante, los rusos han inundado el mercado de emisiones europeo con su "superávit de EUA". Se ha planteado como solución a la sobreoferta de EUA, que se le aplique un factor de descuento del 30%, 50% u 80% a los créditos que generan el "superávit de EUA rusos".

Cabe recordar que el grueso de la generación eléctrica en Europa, EUA, Rusia, China, India y Japón, está soportada en centrales de generación térmica que necesariamente queman combustibles fósiles como carbón o gas natural.

## 2.2. Descripción del sistema "Cap and Trade" que crea los mercados de emisiones de CO<sub>2</sub>

En el protocolo de Kyoto, artículo 17, fue introducido el "Emission Trading System", ETS [11], o sistema transaccional, para crear un mercado de emisiones que funcione de manera similar a cómo funcionan

los mercados de commodities o productos básicos no procesados en Estados Unidos, Europa y Japón. El ETS se basó en la exitosa experiencia de los americanos, quienes ya habían creado un mercado de emisiones para combatir la lluvia ácida -causada por la emisión de gases contaminantes- al comienzo de los 90. El mercado de emisiones americano consistió en crear “permisos de polución”, los cuales eran transables. Dichos permisos de polución buscaban reducir a bajo costo la emisión de plomo y dióxido de sulfuro al medio ambiente; este mercado fue creado por el Congreso mediante el “US Clean Air Act Amendments of 1990”. El mercado de emisiones contra la lluvia ácida se estima que ayudó a reducir los costos de emisión por encima del 50% [12].

La siguiente es una descripción detallada de cómo los ETS trabajan:

Dado que los permisos están limitados a una cantidad menor que la cantidad de gas que normalmente se emite, el derecho a emitir se convierte en una mercancía valiosa. Si se permite el comercio de permisos, entonces, un precio de mercado será establecido. Aquellos que deseen emitir los gases especificados más allá de los niveles permitidos deben o bien reducir sus emisiones (mediante la implementación de cambios tecnológicos) o comprar permisos para emitir. Los contaminadores capaces de reducir sus emisiones a un relativo bajo costo, lo harán en lugar de comprar permisos para contaminar. Los contaminadores que se enfrenten a mayores costos de reducción tenderán a comprar permisos para satisfacer los requerimientos gubernamentales. De esta manera, las reducciones en las emisiones son hechas por los contaminadores que puedan hacerlo al menor costo posible, siendo compensados por los contaminadores que enfrentan mayores costos de reducción [12].

El sistema “Cap and Trade” para CO<sub>2</sub> de los europeos es una versión mejorada del sistema creado por los americanos para combatir la lluvia ácida. Este sistema está basado en estimular la reducción gradual de emisiones de gases con efecto invernadero. Las metas concretas para el año 2020 son 21% más bajas que el año 2005 y para el año 2030, más ambiciosas aún, 43% más bajas que en 2005. Desde el año 2013, 11000 empresas y centrales de generación térmica están obligadas a reducir sus emisiones 1.47% cada año hasta lograr un 21% de reducción total para la Fase III en el año 2020.

El sistema lo conforman los 28 miembros de la Unión Europea más Noruega, Lichestein e Islandia. Los gobiernos emiten cada año un número determinado de EUA, permisos para contaminar. Durante la Fase II, el 100% de los EUA emitidos por los gobiernos era gratuito -las emisiones hechas por los gobiernos se consideran emisiones primarias-; en la Fase III, se introdujeron cambios significativos para hacer el sistema más exigente. A partir del año 2013, el 80% de los EUA para la industria manufacturera fueron emitidos de manera gratuita; sin embargo, este porcentaje irá decreciendo gradualmente hasta alcanzar un 30% en el año 2020. El 20% restante de los EUA correspondientes al 2013 fueron emitidos por el mecanismo de subasta pública, en el cual cada año se incrementará gradualmente el porcentaje de EUA subastados hasta alcanzar 70% en 2020, es decir, las industrias comenzaron a “pagar por contaminar”. Adicionalmente, las centrales térmicas de generación están en otra categoría y durante toda la Fase III están obligadas a comprar mediante mecanismo de subasta el 100% de los EUA que requieran, de acuerdo a su respectivo nivel de contaminación anual [11].

La compra de los EUA por parte de las centrales térmicas de generación, sumado al hecho de que cada año reciben un menor número de EUA, puede obligar a dichas plantas a implementar cambios tecnológicos o migrar hacia sistemas de generación menos contaminantes -en términos de emisiones-. Por ejemplo, si el “switching cost” por cambiar de carbón a gas natural como combustible para la central es superior a lo que cueste comprar los EUA, la decisión es simple: se “paga por contaminar”. El “switchig cost” está directamente relacionado con el precio de los EUA en el mercado de emisiones europeo, es decir, a cuánto se están transando los EUA. Si el precio de los EUA es alto, supóngase por encima de 20 euros, comienza a verse interesante hacer los cambios tecnológicos necesarios para cambiar de carbón a gas natural. Otra alternativa para los grandes generadores sería migrar hacia tecnologías limpias de generación, donde no tendrían que comprar EUA, como son la eólica, la solar fotovoltaica o la geotérmica [11].

### 2.2.1. *Problemas para la implementación del “Sistema Cap and Trade”*

El sistema “Cap and Trade” ha tenido problemas causados principalmente por la sobreoferta de EUA en las Fases II y III. La sobreoferta ha colapsado ar-

tificialmente la formación de precios de transacción. Los bajos precios de los EUA, a su vez, han desestimulado a las industrias manufactureras y de generación eléctrica a adoptar los cambios tecnológicos requeridos para disminuir drásticamente y, en períodos de tiempo cortos, la enorme cantidad de emisiones de gases de efecto invernadero. Surge entonces la pregunta ¿Por qué simplemente no se cambia el sistema “Cap and Trade” por un sistema impositivo que penalice las emisiones de gases con efecto invernadero que sea más eficaz en lograr reducciones ambiciosas y que estimule la migración hacia tecnologías de generación eléctrica renovables?

La respuesta a esta pregunta, formulada en el párrafo anterior, nos indicará cuál es el sistema ideal a adoptar en Colombia, en caso tal de que el gobierno decida estimular la reducción de emisiones de CO<sub>2</sub> por parte de la industria manufacturera y principalmente para las centrales térmicas. En este punto es fundamental resaltar que el sistema ideal para estimular la reducción de emisiones de CO<sub>2</sub> será el que se utilice en este trabajo, para poder valorar el costo de dichas emisiones. Luego, el costo por tonelada de CO<sub>2</sub> emitida será incorporado como un costo variable adicional del proceso de generación eléctrica, por parte de las centrales térmicas en Colombia. Incorporar el costo por tonelada de CO<sub>2</sub> emitida dentro los costos de generación eléctrica de las centrales térmicas en Colombia permite aproximar aún más sus respectivas estructuras de costos a estándares internacionales.

Para dar respuesta a la pregunta previamente formulada, primero se analizan las características del sistema “Cap and Trade”. Se puede afirmar que este sistema premia financieramente los esfuerzos que las industrias altamente contaminantes y el sector de generación eléctrica hagan para reducir sus emisiones anuales de CO<sub>2</sub> por debajo de las metas inicialmente impuestas. Esto quiere decir que la sobreredución de emisiones crea un estímulo financiero, ya que la industria que realice los cambios a tecnologías limpias de producción o generación puede acumular los EUA previamente asignados y vender o transar los EUA que le sobren año tras año, lo cual crea nuevos ingresos para esas industrias. También se presenta el caso de las industrias que no adopten tecnologías más limpias de producción o generación eléctrica: estas industrias siempre estarán “cortas” respecto de la cantidad de EUA asignados anualmente. Por lo tanto, las industrias más contaminantes se verán obligadas a comprar en el

mercado de emisiones su faltante de EUA, el cual será el “sobrante” de las industrias más limpias que estarán “largas en EUA”.

Ahora se analizan las consecuencias de adoptar un sistema impositivo que penalice fiscalmente las emisiones de CO<sub>2</sub>. Se observa que esta política, al no poseer un incentivo financiero, genera un efecto contrario al buscado, o sea, desestimula las industrias para lograr el cumplimiento de las metas impuestas en términos de reducción de emisiones de CO<sub>2</sub>. Una política de penalización fiscal a las emisiones de gases con efecto invernadero crea un efecto psicológico consistente en que accionistas y gerentes del sector de generación eléctrica vean la norma como una persecución a su actividad económica por el hecho de ser contaminante, en lugar de ver esta política como una medida que busca la incorporación de tecnologías más limpias y eficientes que reduzcan los niveles de emisiones. Por el contrario, el sistema “Cap and Trade”, dado que posee un estímulo financiero por superar las metas en reducción de emisiones, se convierte en un estímulo permanente para alcanzar, año tras año, metas más ambiciosas en términos de reducción de emisiones.

Con base en el análisis presentado, se observa que el sistema “Cap and Trade” utilizado por los europeos tiene ventajas sobre un sistema de penalización fiscal a las emisiones de gases con efecto invernadero. La respuesta a la pregunta formulada es que el sistema “Cap and Trade” es más eficaz en lograr reducciones significativas de las emisiones de CO<sub>2</sub>, a pesar de que el precio de los EUA transados se encuentre actualmente demasiado bajo. En conclusión, en este trabajo se adoptará el precio de los EUA para valorar el costo de cada tonelada de CO<sub>2</sub> emitida por centrales térmicas de generación eléctrica. Dicho costo será incorporado como un costo variable adicional e inherente a las centrales térmicas de generación eléctrica en Colombia.

### 2.3. Índices que conforman el “Sistema Cap and Trade”

El “Sistema Cap and Trade” está conformado por dos familias de índices o instrumentos financieros. La primera familia la conforman todos los títulos CER, mientras que la segunda familia corresponde a los títulos EUA.

### 2.3.1. Promedio de precios históricos del instrumento "Futuro CER"

El comportamiento histórico de los CER se ha visto afectado por la sobreoferta de títulos en el mercado de emisiones europeo. Probablemente la actual recesión en la Eurozona ha movido los cimientos del mercado de emisiones europeo. Lo anterior se traduce en que los gobiernos Europeos destinan cada vez menos presupuesto para financiar desarrollos limpios por fuera de la Eurozona. Adicionalmente, los flujos de capital de inversionistas privados tienden a irse hacia el desarrollo de tecnologías innovadoras en términos de eficiencia y menos dependientes de combustibles fósiles, es decir, el capital privado prefiere apostarle a estas tecnologías innovadoras sin incurrir en los trámites burocráticos que los CDM requieren [13]. En síntesis, la suma de todos los factores previamente descritos hace que la transacción de todos los títulos CER se haya deprimido ostensiblemente. En la Figura 3, se observa el colapso del precio de los títulos CER que se transan a diario en la plataforma del Intercontinental Exchange; la línea gris clara muestra que dichos títulos en la actualidad se acercan a 0 Euros.

### 2.3.2. Promedio de precios históricos del instrumento "Futuro EUA"

En la Figura 4, se encuentra la gráfica correspondiente a la serie de precios diarios de transacción del instrumento "Futuro EUA", los cuales se encuentran disponibles en el "Emissions Index" del Intercontinental Exchange [15].

Al observar la Figura 4, se observa que el instrumento tuvo sus más altos precios de transacción al comienzo de la fase I. Luego se desplomó debido a que las empresas en dicho período no se vieron obligadas a disminuir sus emisiones. Dado el hecho de que los títulos emitidos en la fase I expiraban con la terminación de esta, se vio un repunte en los precios al comienzo de la fase II. Una vez los títulos emitidos por la Federación Rusa inundan el mercado en la fase II, los precios se desploman de nuevo. En la fase III, los precios tocan mínimos históricos por debajo de 4 euros; sin embargo, rebotan y en la actualidad se transan cerca a los 7 euros.

### 2.3.3. Un nuevo enfoque para valorar el costo de las emisiones de CO<sub>2</sub> -concepto SCC-

El costo social del carbono (o SCC por sus siglas en Inglés) es un concepto más avanzado para valorar las emisiones de CO<sub>2</sub> que los EUA de los europeos. Este concepto es teórico y NO involucra instrumentos financieros o futuros de emisiones que se puedan transar en el Intercontinental Exchange (ICE). Los SCC, aunque son elementos de valoración teóricos no transables, pueden de hecho valorar de una manera más aproximada y real, el costo de las emisiones de CO<sub>2</sub>.

En la Figura 5, se correlaciona el incremento del promedio global del temperatura con el incremento en la concentración de CO<sub>2</sub>, medida en partículas por millón (ppm) en la atmósfera. En el eje vertical izquierdo se observa la temperatura global en Fahrenheit (F); el eje vertical derecho representa la concentración de CO<sub>2</sub> en partículas por millón (ppm) y el eje horizontal representa la escala de tiempo en años. Las líneas azules representan los años en los cuales la temperatura superficial de la tierra ha estado por debajo del promedio histórico. Las líneas rojas representan los años en los cuales la temperatura superficial de la tierra ha estado por encima del promedio histórico. La línea negra representa la medición anual de concentración de CO<sub>2</sub> en ppm. Se puede observar la correlación entre el incremento de la concentración de ppm y el incremento de la temperatura medida en la superficie de la tierra sobre el promedio global histórico, según Interagency Working Group on Social Cost of Carbon [8].

Primero, los SCC no están afectados por las leyes de oferta y demanda que rigen el mercado. Segundo, los SCC buscan valorar el daño económico asociado al incremento de la temperatura promedio en la superficie terrestre, debido al incremento en la concentración de partículas por millón o ppm de CO<sub>2</sub> en la atmósfera, que es causada por la emisión de CO<sub>2</sub> al quemar combustibles fósiles. Esta valoración del daño económico se hace al convertir las emisiones proyectadas de CO<sub>2</sub> en incrementos de temperatura, que ocasiona impactos negativos socioeconómicos en los países y en el crecimiento del PIB de cada país.

El SCC aparece en un documento de soporte técnico para análisis del impacto regulatorio elaborado para la casa blanca, por un amplio grupo interinstitucional que involucra diferentes agencias, departamen-



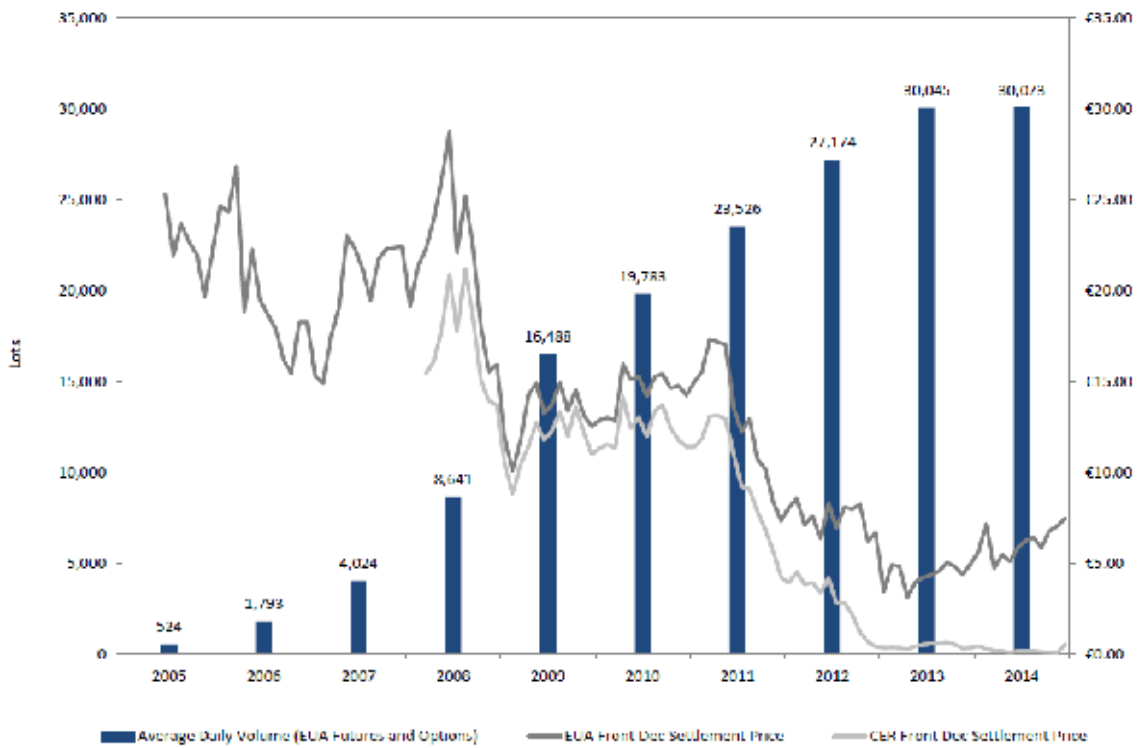


Fig. 3. Precio promedio diario Títulos CER y EUA, tomado de [14]



Fig. 4. Precios diarios del instrumento "Futuro EUA". Fuente: Elaboración propia con datos tomado de [15]

tos y secretarías del gobierno federal de los Estados Unidos. El propósito de los SCC es “permitirle a las agencias incorporar los beneficios sociales de la reducción de dióxido de carbono (CO<sub>2</sub>) en el análisis de costo-beneficio de las medidas normativas que tengan impactos pequeños, o “marginales” [8], en las emisiones globales acumuladas de CO<sub>2</sub>”. Otra manera de entender los SCC es “un estimado de los daños monetizados, asociados con un aumento incremental de las emisiones de CO<sub>2</sub> en un año dado” [8]; dicho estimado se proyecta en USD por tonelada métrica de CO<sub>2</sub> emitida.

El SCC puede ser visto entonces como un costo asociado a los daños causados por el cambio climático debido a las emisiones de CO<sub>2</sub>. También puede ser visto como un beneficio por la reducción de emisiones de CO<sub>2</sub>. De hecho, al valorar el costo de dichos daños, lo que se busca es concientizar a los tomadores de decisiones para que adopten diferentes actitudes y cambios tecnológicos que permitan reducir a nivel global las emisiones de CO<sub>2</sub>. Los valores de los SCC dados desde el 2010 hasta el año 2050 resultan de tres diferentes modelos de valoración integrada, donde los datos desde 2010 hasta 2013 son los estimados más recientes y reales, mientras los datos de 2014 a 2050 son datos proyectados. La más reciente actualización del documento corresponde a noviembre de 2013. Para profundizar sobre el SCC, favor leer [7]. Los modelos se corren para diferentes escenarios y tasas de descuento, donde la tasa de descuento más realista es del 3% y se hace con dólares constantes de 2007. La valoración inicia en 2010 con un costo de 32 USD por ton de CO<sub>2</sub> emitida, mientras que para 2014, dicho costo se ha incrementado a 36 USD por tonelada de CO<sub>2</sub> emitida.

#### 2.3.4. Precio histórico del “concepto SCC”

Los precios históricos de los SCC son presentados anuales en el documento de [7] y aparecen valorados a partir del año 2010. Con el ánimo de hacer el ejercicio académico, en este trabajo se extrapolan valores para obtener los precios en los años 2006, 2007, 2008 y 2009 -que no aparecen en dicho documento-; en general, se puede afirmar que los valores de SCC estimados por los modelos aumentan a una rata de 1 USD/año. En la Tabla 1 se presentan los valores utilizados en este trabajo.

### 3. RESULTADOS ESPERADOS

Se hace pertinente presentar una síntesis sobre el uso de los diferentes instrumentos financieros que surgen a partir de los CDM y su posible aplicación para el desarrollo de proyectos en el sector eléctrico latinoamericano. Es importante resaltar que dado el desplome de los títulos CER y las restricciones impuestas para su emisión en la fase III, se tiene una gran incertidumbre frente a la permanencia de estos títulos en el futuro cercano.

#### 3.1. Aplicación de los CER en el sector eléctrico colombiano

En caso tal de que los CER subsistan y se recupere su precio, estos títulos pueden ser utilizados para financiar parte de la fase constructiva de nuevos proyectos de generación eléctrica con fuentes renovables, es decir, nuevos proyectos de generación eléctrica que hagan uso de fuentes hidráulicas o eólicas podrían aplicar específicamente por títulos CER. Los títulos ACM002 aplicarían para la financiación de hidroeléctricas con capacidad mayor de 20MW, mientras que los AMS-I.D. aplicarían para la financiación de hidroeléctricas con capacidad menor de 20MW. Una alternativa para incrementar la diversificación de la matriz eléctrica mediante fuentes renovables es la generación eólica. Los títulos AMS-I.B. aplicarían para parte de la financiación de proyectos eólicos [6].

Independiente del futuro de los CER, se debe resaltar que nuevos proyectos de generación eólica en Colombia resultan atractivos a los ojos de inversionistas, debido a dos razones. Primero, la última tecnología de aerogeneradores, como el Nórdex N117/3000 [16], presenta un notable incremento en su factor de planta. Este incremento hace su operación rentable y competitiva al comparar sus costos de generación con los costos de las plantas térmicas a gas o carbón [1]. El segundo elemento que hace aún más competitivos los costos de la generación eólica son los beneficios de la ley colombiana 1715 de 2014, que busca incentivar el uso de nuevas tecnologías renovables. La ley 1715 incluye beneficios como la exención de aranceles para la importación de equipos, exclusión del pago de IVA de dichos equipos, además de una reducción de renta “por los 5 años siguientes al año gravable en que hayan realizado la inversión, del cincuenta por ciento (50%) del valor total de la inversión realizada” [17]

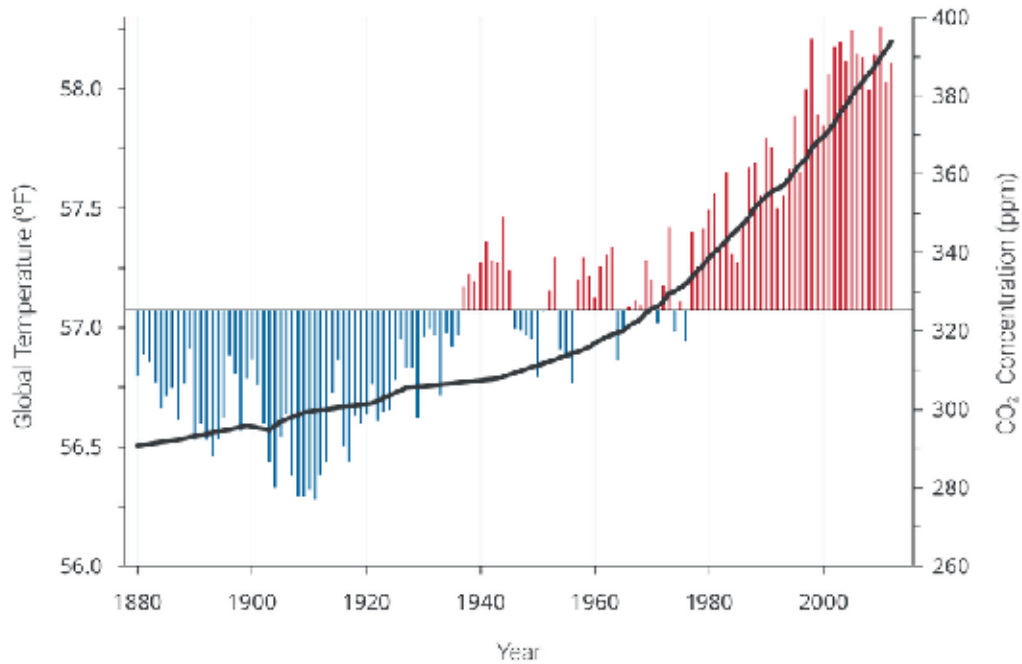


Fig. 5. Concentración de CO2 en (ppm), tomada de <http://www.whitehouse.gov/climate-change>

Tabla 1. Precios históricos del “concepto SCC”. Fuente: Elaboración propia [1].

Año	SCC (USD / ton de CO2)
2006	28
2007	29
2008	30
2009	31
2010	32
2011	33
2012	34
2013	35

DATOS EXTRAPOLADOS

DATOS DE DOCUMENTO [7]

### 3.2. Aplicación del instrumento “Futuro EUA”

Básicamente, los EUA son instrumentos apropiados para valorar el costo de las emisiones de CO2 en centrales térmicas de generación y desincentivar el uso de estas tecnologías contaminantes. Los EUA serían los instrumentos a elegir por los países de la región, en caso tal de que se busque desincentivar el uso de tecnologías contaminantes y simultáneamente incentivar el uso de tecnologías no contaminantes como la eólica. Lo anterior se explica en el hecho de que los EUA imponen un sobre costo significativo en los costos variables de operación para los generadores térmicos.

### 3.3. Aplicación del “concepto SCC”

Al igual que los EUA, el concepto SCC sirve para valorar el costo de las emisiones de CO2 en las centrales térmicas, además de desincentivar su uso. El concepto SCC no está afectado por la volatilidad que puede tener un instrumento financiero como el futuro EUA -debido a las leyes de oferta y demanda-. En principio, el concepto SCC puede tener una valoración más real del costo de las emisiones de CO2, si se compara con el futuro EUA.

**Tabla 2.** Costos promedio de generación, considerando EUA, SCC; período (2006-2013) Fuente: Elaboración propia.

Tipo de Tecnología	Costo promedio (2006-2013) COP	Costo con EUA (COP)	Costo con SCC (COP)
Hidráulica (H)	104,28	104,28	104,28
Térmica carbón (TC)	104,88	145,21	173,02
Térmica Gas (TG)	126,51	146,95	161,05
Eólica (E) ; (N80/2500)	172,01	172,01	172,01
Eólica (E) ; (N117/3000)	104,52	104,52	104,52

### 3.4. Costos de generación por tecnología, considerando los futuros EUA y el concepto SCC

Una manera de visualizar el impacto sobre los costos de generación que causa valorar el costo de las emisiones de CO<sub>2</sub> consiste en comparar los costos de diferentes tecnologías de generación, haciendo uso de los futuros EUA y del concepto SCC. En la Tabla 2, se presentan los costos promedio de generación de las tecnologías hidráulica (H), térmica carbón (TC), térmica a gas (TG) y eólica (E), para el período comprendido entre los años 2006 a 2013, en pesos colombianos (\$ o COP). Con respecto a los costos de la energía eólica, es importante mencionar que son valores calculados con el régimen de vientos en el intervalo de tiempo mencionado (2006-2013), medido en el sitio Puerto Bolívar, Guajira, Colombia. Como resultado de estos cálculos se presentan dos valores teóricos de costos: el primero es el costo de generación para una turbina NORDEX N80/2500; El segundo costo eólico es el calculado para una turbina NORDEX N117/3000, la cual pertenece a una nueva generación de turbinas con mayor eficiencia en comparación con las anteriores. Se observa claramente que con esta nueva turbina, el costo de generación sería altamente competitivo en comparación con las otras tecnologías.

Se observa que la energía hidroeléctrica es la de menor costo, los costos de las térmicas en la primera columna son competitivos. Sin embargo, al incluirle a estos costos lo referente a las emisiones de gases de efecto invernadero (EUA y SCC), se observa cómo los costos suben dramáticamente, pasando a ser menos competitivas aún que la energía eólica.

### 4. CONCLUSIONES

En general, se concluye que la valoración definida por la UNFCCC no es suficiente para identificar e impulsar todo el potencial de energía renovable posible. Se recomienda tomar otros indicadores como el SCC.

Se ha encontrado gran volatilidad en los indicadores derivados de los CDM; gran parte de esta es debido a la dinámica del mercado, la cual a está influenciada por condicionamientos políticos y macroeconómicos de los países desarrollados y de los países en desarrollo [3].

Dado que actualmente los precios de los CER se encuentran cercanos a cero euros, NO se recomienda utilizar estos instrumentos como fuente de financiación alternativa para la fase constructiva de proyectos de generación con fuentes renovables. Solo se recomienda hacer uso de los CER en caso tal de que su precio vuelva a estar por encima de los 7 Euros.

Los futuros EUA y el concepto SCC son ambos apropiados para valorar el costo de las emisiones de CO<sub>2</sub> en plantas de generación térmica. Su uso es válido en caso tal de que los países decidan incentivar el uso de tecnologías de generación con fuentes renovables y simultáneamente desincentivar el uso de fuentes contaminantes.

Con respecto al caso colombiano, se evidencia que, dependiendo de la valoración que se escoja para las emisiones, las tecnologías renovables como la eólica podrían tener más oportunidad de penetrar en el mercado. Adicionalmente se observa que si se decide valorar el costo de las emisiones de CO<sub>2</sub> y se carga este costo variable a los costos totales para las plantas térmicas a carbón y térmicas a gas, estas pierden competitividad frente a las tecnologías renovables [1].

## 5. LISTA DE SÍMBOLOS Y ABREVIATURAS

Abreviatura / Símbolo	Término
CO <sub>2</sub>	Gas, Dióxido de carbono
OCDE	Organización para la Co-operación y el Desarrollo Económico
MDL	Mecanismos de Desarrollo Limpio
CDM	Clean Developmet Mechanism
CER	Certificados de Reducción de Emisiones
UE	Unión Europea
UNFCCC	United Nations Framework Convention on Climate Change
IET	International Emission Trading
JI	Joint Implementation
EU ETS	European Union Emission Trading Scheme
EUA	European Union Allowances
ERU	Emission Reduction Unit
MW	Mega Watt
ACM002	Título CER específico
AMS-I.D.	Título CER específico
AMS-I.B.	Título CER específico
SCC	Costo Social del Carbono
ICE	Intercontinental Exchange
F	Fahrenheit
ppm	partículas por millón
USD	Dólar Americano
\$ o COP	Peso Colombiano
H	Hidráulica
TC	Térmica a carbón
TG	Térmica a Gas
E	Eólica

## REFERENCIAS

- [1] M. A. Berrío Monsalve, "Diversificación de la canasta eléctrica colombiana considerando costos de emisiones de CO<sub>2</sub>," Universidad Nacional de Colombia, 2014.
- [2] C. González, "Definición de la composición en las fuentes hidráulica y eólica para la generación de energía eléctrica en el contexto colombiano aplicando la teoría de portafolio," 2012. Disponible: [http://www.bdigital.unal.edu.co/7033/1/91276818.\\_2012.pdf](http://www.bdigital.unal.edu.co/7033/1/91276818._2012.pdf).
- [3] A. K. Dhamija, "Financial risk & opportunities of Carbon finance: Price determinants and volatility estimation of EUA & CER," 2011. Disponible: [http://akdhamija.webs.com/my\\_docs/rp.pdf](http://akdhamija.webs.com/my_docs/rp.pdf).
- [4] OECD, "OECD Economic Surveys: Colombia 2015 | OECD READ edition," 2015. Disponible: [http://www.keepeek.com/Digital-Asset-Management/oecd/economics/oecd-economic-surveys-colombia-2015\\_eco\\_surveys-col-2015-en#page82](http://www.keepeek.com/Digital-Asset-Management/oecd/economics/oecd-economic-surveys-colombia-2015_eco_surveys-col-2015-en#page82). [Accessed: 22-Apr-2015].
- [5] E. Duque, "Sistemas de bono de carbono como fuente de restauración del capital natural en proyectos hidroeléctricos," 2013. Disponible: <http://www.bdigital.unal.edu.co/20391/1/71290657.2014.pdf>.
- [6] United Nations Framework Convention on Climate Change, *Clean Development Mechanism* (CDM), vol. 2014, November, 2014.
- [7] U.S. Environmental Protection Agency (EPA), "Technical Support Document: Social cost of carbon for regulatory impact analysis under Executive Order 12866," 2013. Disponible: <http://www.epa.gov/otaq/climate/regulations/scc-tsd.pdf>.
- [8] U. S. G. Interagency Working Group on Social Cost of Carbon, "Technical Support Document : Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866," 2010. Disponible: <http://www.epa.gov/oms/climate/regulations/scc-tsd.pdf>.
- [9] United Nations, "Kyoto Protocol to the United Nations Framework Convention on Climate Change," 1998. Disponible: <http://unfccc.int/resource/docs/convkp/kpeng.pdf>.
- [10] United Nations Framework Convention on Climate Change, "Glossary of CDM terms," 2001. Disponible: [http://cdm.unfccc.int/Reference/Guidclarif/glos\\_CDM.pdf](http://cdm.unfccc.int/Reference/Guidclarif/glos_CDM.pdf).
- [11] European Commission, "The EU Emissions Trading System (EU ETS)," 2014. Disponible: [http://ec.europa.eu/clima/policies/ets/index\\_en.htm](http://ec.europa.eu/clima/policies/ets/index_en.htm).
- [12] C. Damro and P. L. Méndez, "Emissions trading at Kyoto: from EU resistance to Union innovation," *Environmental Politics*, 2003. Disponible: <http://dx.doi.org/10.1080/09644010412331308194>.
- [13] F. Harvey, "Global Carbon Trading System Has 'Essentially Collapsed,'" *The Guardian*, 2012. Disponible: <http://ourworld.unu.edu/en/global-carbon-trading-system-has-essentially-collapsed>.
- [14] Intercontinental Exchange, "The Emissions Market ICE Futures Europe," 2014. Disponible: [www.theice.com/products/814666/CER-Futures](http://www.theice.com/products/814666/CER-Futures).
- [15] Intercontinental Exchange, "EUA Futures Europe," 2014. Disponible: <https://www.theice.com/products/197/EUA-Futures/data>.
- [16] Nordex, "Wind Turbines," 2014. Disponible: <http://www.nordex-online.com/en>.
- [17] Congreso de Colombia, "Ley N° 1715 Del 13 de Mayo de 2014", 2014. Disponible: [http://wsp.presidencia.gov.co/Normativa/Leyes/Documents/LEY\\_1715\\_DEL\\_13\\_DE\\_MAYO\\_DE\\_2014.pdf](http://wsp.presidencia.gov.co/Normativa/Leyes/Documents/LEY_1715_DEL_13_DE_MAYO_DE_2014.pdf).



# CrawNet: Crawler de Recursos Multimedia para la Web Superficial y Oculta

## CrawNet: Multimedia Crawler Resources for Both Surface and Hidden Web

### Alicia Martínez-Rebollar, Ph.D

*Departamento en Ciencias de la Computación, Centro Nacional de Investigación Científica y Tecnológica Cuernavaca, Morelos, México  
amartinez@cenidet.edu.mx*

### Fernando Pech-May, MSc

*Departamento en Ciencias de la Computación, Centro Nacional de Investigación Científica y Tecnológica Cuernavaca, Morelos, México  
fpech@cenidet.edu.mx*

### Hugo Estrada-Esquivel, Ph.D

*Gerencia de Desarrollo de Nuevos Productos y Servicios, Fondo de información y documentación para la Industria, INFOTEC Ciudad de México DF, México  
hugo.estrada@infotec.com.mx*

### Eduardo Pedroza-Landa, MSc

*Departamento en Ciencias de la Computación, Centro Nacional de Investigación Científica y Tecnológica Cuernavaca, Morelos, México  
eduardopl11c@cenidet.edu.mx*

(Recibido el 20-10-2014. Aprobado el 20-12-2014)

**Resumen.** La web es la fuente de información de mayor uso en el ámbito académico, científico e industrial. Su crecimiento explosivo ha generado billones de páginas con información, las cuales se categorizan como web superficial, integrada por páginas estáticas que pueden ser indexadas; y web oculta, accesibles a través de formularios de búsqueda. En este artículo, se presenta el desarrollo de un crawler que permite realizar búsquedas, consultas y análisis de información en la web superficial y oculta en dominios específicos de la web.

**Palabras clave:** Crawler; web oculta; web superficial.

**Abstract.** The web is the most used information source in both academic, scientific and industry forums. Its explosive growth has generated billions of pages with information which may be categorized as surface web, composed of static pages that are indexed into a hidden web, accessible through search templates. This paper presents the development of a crawler that allows searching, queries, and analysis of information in the surface web and hidden in specific domains of the web.

**Keywords:** Crawler; hidden web; surface web.

**Citación de artículo, estilo IEEE:**

A. Martínez-Rebollar, F. Pech-May, H. Estrada-Esquivel, E. Pedroza-Landa, "CrawNet: Crawler de Recursos Multimedia para la Web Superficial y Oculta", *Lámpsakos*, N° 13, pp. 39-50, 2015.

DOI: <http://dx.doi.org/10.21501/21454086.1365>

## 1. INTRODUCCIÓN

La World Wide Web ha tenido un crecimiento explosivo que ha generado grandes cantidades de información disponible en la web; actualmente representa la fuente de información más utilizada en el sector académico, científico e industrial.

Los motores de búsqueda son el mecanismo más utilizado para consultar la información contenida en la web. Estos motores coleccionan, de forma masiva, páginas que adquieren con la ayuda de rastreadores web (crawler), que atraviesan la web siguiendo los hipervínculos y almacenan las páginas en una gran base de datos para posteriormente indexarlas. De esta manera, cuando un usuario realiza una petición a un motor de búsqueda, este consulta su base de datos en la cual se han almacenado páginas indexadas a través de un crawler específico [1].

Un estudio realizado por BrightPlanet [2] demuestra que la mayoría de los motores de búsqueda solo acceden a información localizada en la web superficial, compuesta por documentos estáticos accesibles a través de ligas de navegación. Sin embargo, existe una gran cantidad de información en la denominada web oculta, solamente accesibles a través de formularios de búsqueda. Hoy en día, se hace difícil el acceso a información de la web oculta debido a que está constituida por bases de datos, formularios, contenido script, etc. que requieren la intervención del usuario [3].

En este artículo se presenta el desarrollo de un crawler, denominado CrawNet, de recursos multimedia para la web superficial y oculta, que permite realizar búsquedas, consultas y análisis de información en dominios específicos por medio de patrones de expresiones regulares que buscan relaciones semánticas entre conceptos.

El documento está organizado de la siguiente manera: en la sección 2, se describe el marco teórico; en la sección 3, se presenta la arquitectura del crawler para la web superficial y oculta; la sección 4 presenta los detalles de implementación; la sección 5 expone los resultados de la evaluación del Crawler; finalmente, en la sección 6 se presentan las conclusiones y trabajos futuros de esta investigación.

## 2. MARCO TEÓRICO Y TRABAJOS RELACIONADOS

### 2.1. Web superficial y oculta

La web superficial está constituida por páginas estáticas y que pueden ser indexadas por los motores de búsqueda tradicionales [4]. Según Fernández & Pardo [5], casi el 85% de los usuarios navegan en esta web. Se estima que el tamaño de la web oculta supera por mucho a la web superficial [6], [7]. En el año 2000, la web superficial era de 7,500 TeraBytes mientras que 167 TeraBytes eran de web oculta [8], [5].

La web oculta, también denominada web profunda o invisible, es aquella que contiene información que no es posible tener acceso a ella mediante un motor de búsqueda tradicional, ya que son casi inexplorados o no indexados [9] debido a que hace referencia a un contenido detrás de formularios HTML, por lo que es necesario rellenar formularios para realizar una consulta específica, autenticación de usuario [10] o la existencia de sitios dinámicos que utilizan scripts [11].

La web oculta contiene información considerada de gran valor [12]. Existen autores que proponen y muestran datos de estimaciones estadísticas acerca del tamaño actual de dicha información [13]. Sherman [14] propone una clasificación de la web oculta basándose en su contenido; dicha clasificación contempla 4 tipos de contenido: la web opaca, la web privada, la web propietaria y la verdadera web invisible.

### 2.2. Crawlers

Los crawlers son pequeños programas utilizados por los motores de búsqueda para la localización, proceso de análisis y almacenamiento de la información. Se clasifican en dos tipos: crawlers de la web superficial y crawlers de la web oculta. A continuación se describe cada una de estas categorías.

- a. Crawlers para la web superficial. Son programas capaces de procesar y analizar la web. Se dedican a recorrer la web superficial localizando páginas estáticas para posteriormente extraer texto HTML e indexarlas [15].



b. Crawlers para la web oculta. Están enfocados en indexar contenido de páginas dinámicas, lo cual no pueden realizar los motores de búsqueda tradicionales, ya que la mayoría trabajan sobre páginas estáticas; sin embargo, carecen de efectividad en el manejo de información debido al gran tamaño de la web oculta; además, el acceso a las bases de datos depende de los datos introducidos en un formulario [16].

### 2.3. Trabajos relacionados

Actualmente se han propuesto diferentes trabajos de investigación que tienen como objetivo acceder a la web oculta [17], [18], [19]. En este artículo nos concentramos en presentar aquellos proyectos que cuentan ya con una herramienta que permite automatizar el proceso de búsqueda de información.

HiWE [20] es un crawler enfocado en la búsqueda de información del dominio de los semiconductores; usa técnicas para la extracción del contenido, tales como el cálculo de la distancia visual y uso de medidas de similitud textual para comparar los campos y los atributos del dominio. Para la extracción de información ocupa la técnica LITE (Layout Based Information Extraction Technique), técnica que utiliza la extracción semántica, descarta todo lo que no sea texto, tales como imágenes, estilos de fuente y estilos de hoja y, finalmente, ayuda a seleccionar el candidato más apto para asociar una etiqueta con el elemento.

Google's Deep-web Crawl [21] tiene como objetivo indexar la mayor cantidad posible de páginas a partir de millones de formularios HTML. Permite acceder a la web oculta utilizando algoritmos para seleccionar valores de entrada para formularios que aceptan palabras clave y algoritmos para identificar entradas de formularios que solo aceptan valores de un tipo específico. Esto se realiza identificando los elementos HTML existentes en los formularios web (cajas de texto y menús de selección) e ignorando aquellos formularios que contienen información personal y áreas de texto para comentarios de usuarios. Su principal desventaja es que para identificar y enviar datos a formularios, utiliza plantillas de consulta que únicamente toman en consideración aquellos formularios con el método get.

DeepBot [3] es un buscador de la web oculta que recibe como entrada un conjunto de definiciones de dominio; utiliza diferentes heurísticas para identificar automáticamente formularios de consulta relevan-

tes. DeepBot emplea mini buscadores web automatizados; utiliza algoritmos heurísticos que miden la distancia entre los elementos y las etiquetas, de tal forma que puedan asociarse semánticamente los campos existentes. El esquema propuesta contiene los elementos del dominio y las relaciones entre ellos y determina los formularios relevantes y la forma en que los campos del formulario deben ser completados.

## 3. ARQUITECTURA DEL CRAWLER PARA LA WEB SUPERFICIAL Y OCULTA

La metodología propuesta busca solucionar la recuperación de información específica que se encuentra en la web superficial y oculta en un dominio específico. Para esto se desarrollaron dos algoritmos con el propósito de aplicar esta solución; los algoritmos que integran nuestra metodología buscan identificar y recuperar información solicitada. CrawNet recibe como entrada una semilla URL, utilizada como un dominio web, para realizar la búsqueda. En la Figura 1 puede apreciarse la metodología para el desarrollo de CrawNet para la web superficial y oculta, y a continuación se describe cada algoritmo.

### 3.1. Algoritmo para el acceso a la web superficial

El algoritmo permite la recuperación de información y enlaces (URLs) basado en cuatro tipos de búsquedas (general, videos de fútbol, gamecast e información de películas); para lograrlo, recibe la semilla URL para posteriormente recorrer y recuperar todos los enlaces existentes en el contenido HTML. Cada enlace recuperado es comparado y evaluado dependiendo del tipo de búsqueda seleccionado.

Para la evaluación y selección de URLs se declararon patrones de expresiones regulares que contienen palabras clave que se comparan directamente; los URLs que coinciden con los patrones se recuperan y se almacenan en una base de datos (BD). A continuación se describe cada etapa.

*A. Búsqueda de URLs con información solicitada por el usuario.* Busca e identifica URLs de acuerdo al tipo de búsqueda, dando como resultado la recuperación de enlaces candidatos que pueden contener la información solicitada. Cada enlace pasa por filtros definidos por patrones de expresiones regulares, integrados por palabras clave relacionadas semánti-

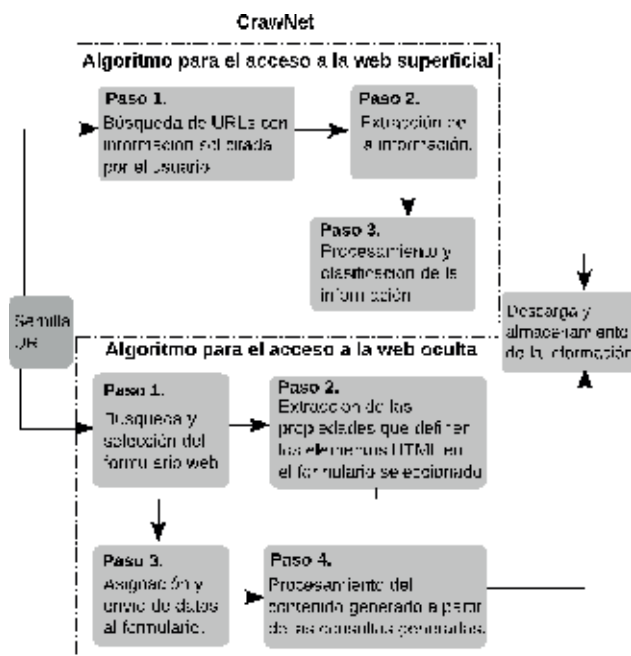


Fig. 1. Metodología CrawNet para la web superficial y oculta.

camente al tipo de búsqueda. Asimismo se encarga de buscar coincidencias entre las palabras clave que integran los patrones de expresiones regulares y las cadenas de texto que conforman un URL. Cuando las palabras clave del patrón coinciden con alguna palabra que se encuentre dentro de un enlace, se selecciona para ser procesado en el siguiente paso.

- B. *Extracción de la información.* Se recibe y procesa el URL seleccionado, el proceso identifica y extrae el enlace si este contiene la información solicitada, basado en los cuatro tipos de búsqueda.
- C. *Procesamiento y clasificación de la información.* Se procesa la información extraída en la etapa anterior. El proceso clasifica y almacena los URLs en una BD. Los URLs y la información recuperada se procesan de acuerdo al tipo de búsqueda utilizada.

### 3.2. Algoritmo para el acceso a la web oculta

Permite identificar páginas web que contienen formularios de búsqueda avanzada y recursos multimedia. Dichos formularios solicitan campos o detalles específicos de un dominio en particular. Los formularios de búsqueda avanzada utilizan un mayor número de campos para describir los elementos buscados y permite que las respuestas sean más cercanas a la solicitada por el usuario. Se tomaron como formularios de búsqueda avanzada aquellos que solicitan

Tabla 1. Patrón para la búsqueda de enlaces y formularios de búsqueda avanzada.

Nombre de patrón	Palabra clave
form_avanzado	Search, advanced-search, advanced_search, advancedsearch, search_advanced, searchadvanced, advanced, adv_search adv-search, advsearch, searchadv, buscar, busquedaavanzada, busqueda_avanzada, busqueda-avanzada, avanzada-busqueda, avanzada_busqueda, busqueda avanzada, avanzada

como mínimo tres campos. CrawNet es capaz de enviar datos a través de dichos formularios para obtener una respuesta y analizarla. Para realizar esto, el algoritmo se basa en cuatro pasos principales.

- A. *Búsqueda y selección del formulario web.* Realiza la búsqueda de parámetros denominados semillas (por ejemplo: título, actor, género y director para películas cinematográficas) y que se encuentran en los formularios de búsqueda avanzada. Para enviar los datos al formulario, es posible usar un solo parámetro, pero se recomienda el uso de más parámetros para la obtención de mejores resultados. La búsqueda en la web se lleva a cabo mediante la búsqueda por expresiones regulares (consiste en identificar y seleccionar un URL como candidato potencial) y por búsqueda de manera exhaustiva (busca directamente en cada web la etiqueta <form>).
- B. *Extracción de las propiedades que definen los elementos HTML en el formulario seleccionado.* Se extraen, mediante un parser, los valores de los atributos de los elementos HTML (form, textbox, combobox y button) los cuales se relacionan semánticamente con los parámetros (título, actor, director y género). De esta manera, se identifica la relación entre las etiquetas y los campos a utilizar para ser llenados por la metodología CrawNet. La extracción de los valores de las propiedades (name, value, action, id) del formulario se realiza mediante el parser Jsoup, dichos valores deben coincidir con el patrón establecido de la Tabla 1.
- C. *Asignación y envío de datos al formulario.* Se realiza el envío de datos al formulario web a través de los lenguajes de navegación HtmlUnit y Selenium, que simulan operaciones de un usuario común sobre páginas web. Estos lenguajes utilizan los valores de las propiedades de los elementos HTML para enviar los datos al formulario;

para lograrlo, se proporciona el valor del atributo del formulario, se le asigna valores semillas a los elementos textbox y/o combobox y finalmente se realiza la consulta con el botón.

D. *Procesamiento del contenido generado a partir de las consultas realizadas.* Se procesa la información obtenida a partir de la consulta realizada por el CrawNet mediante búsquedas de palabras que identifican la búsqueda solicitada dentro del contenido HTML (obtenido como respuesta). Si el elemento es encontrado, el URL es almacenado en la BD. Para analizar la respuesta generada, se lee por completo el documento HTML para aperturar los enlaces que contiene; cada enlace se procesa en busca de valores semillas.

CrawNet fue diseñado para la extracción de recursos multimedia en el dominio cinematográfico, pero adaptable a diferentes dominios. Su arquitectura se encuentra dividida en cinco módulos (ver Figura 2). Los dos primeros módulos se utilizan para ambas búsquedas (superficial y oculta). La entrada principal de esta arquitectura es una semilla URL que representa el dominio web donde ambas búsquedas realizan el proceso de recuperación de información. La búsqueda sobre la web oculta requiere datos adicionales llamados valores semilla (que se utilizan para identificar una semilla). A continuación, se describen cada uno de los módulos que se muestran en la Figura 2.

- *Recuperación de páginas.* Realiza la recuperación y almacenamiento de los URLs contenidos dentro del dominio especificado en la semilla. Como primer paso identifica, obtiene y almacena todos los URLs encontrados en el dominio para posteriormente almacenarlo en una BD. Posteriormente se accede a la BD para recuperar el primer URL e inspeccionarlo a profundidad para la obtención de nuevos enlaces; cada URL es navegado y almacenado en una BD. La búsqueda es finalizada cuando no existe un nuevo URL dentro de las páginas almacenadas.
- *Analizador y clasificador de páginas.* Se encarga de procesar cada enlace seleccionado de acuerdo al tipo de búsqueda. Para la web superficial se utilizan filtros sobre los URLs que identifican y envían los enlaces a las clases respectivas para su procesamiento; posteriormente se almacena en la BD. Para la web oculta se identifican páginas web candidatas a contener formularios de búsqueda avanzada. Esta identificación de páginas candidatas se realiza paralelamente con el primer módulo de dos diferentes maneras: En la primera, se realiza la búsqueda de palabras clave que den indicios de la presencia de un formulario de búsqueda avanzada; para este caso se utiliza la palabra clave “búsqueda avanzada” junto con sus derivaciones y traducciones (search, advanced-search, etc.), las cuales son basadas en las propiedades name, action e id de los formularios; con la segunda manera, se identifican las etiquetas <form> dentro del contenido de las páginas.
- *Analizador de formularios.* Identifica y recupera formularios de búsqueda avanzada junto con los elementos del <form> (textbox, combobox y button) relacionados a un dominio específico; esto se realiza buscando similitud entre los parámetros de dominio, etiquetas y propiedades (name, value e id) de los elementos HTML identificados como candidatos. Se usaron como parámetros de dominio las palabras título, actor, director y género (así como sus derivaciones y traducciones) relacionadas con el tópico de películas cinematográficas. Un formulario de búsqueda es seleccionado solo en caso de haber encontrado al menos tres correspondencias entre elementos del formulario y parámetros de dominio. Finalmente, se identifican valores semilla que servirán para realizar una consulta a través del formulario de búsqueda, obteniendo los valores apropiados para los elementos HTML identificados.

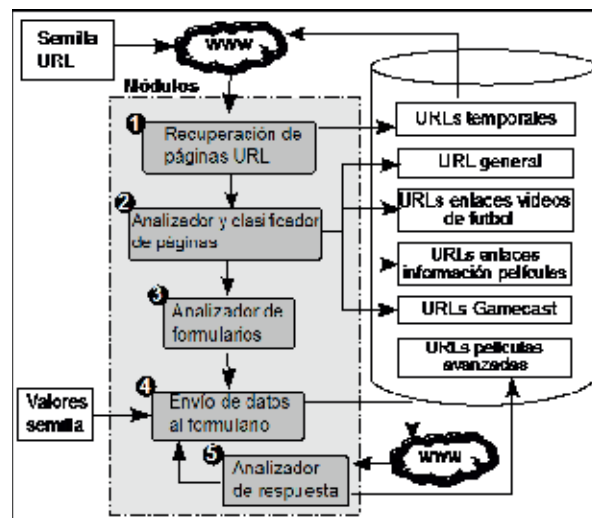


Fig. 2. Panorama general de la arquitectura CrawNet.

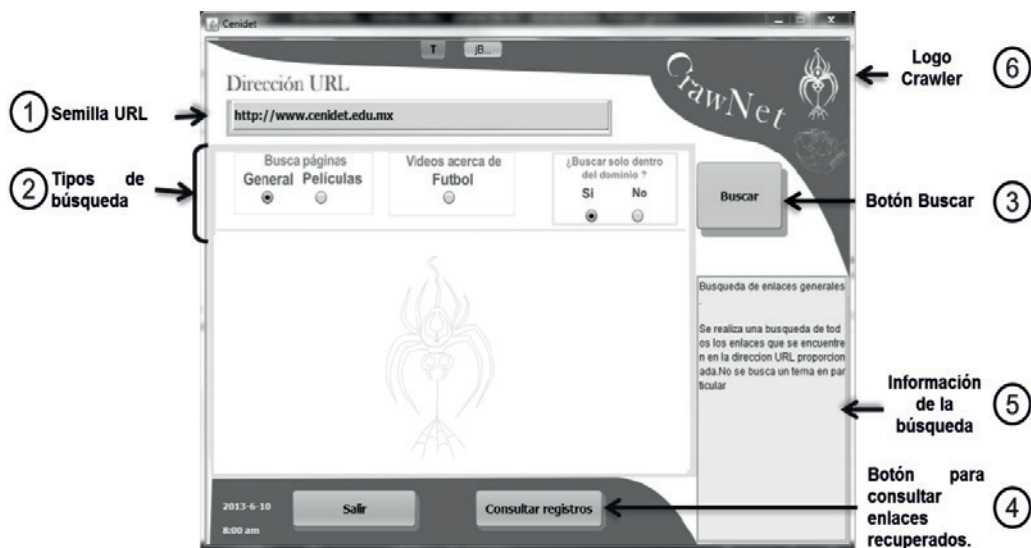


Fig. 3. Interfaz principal de la herramienta CrawlNet.

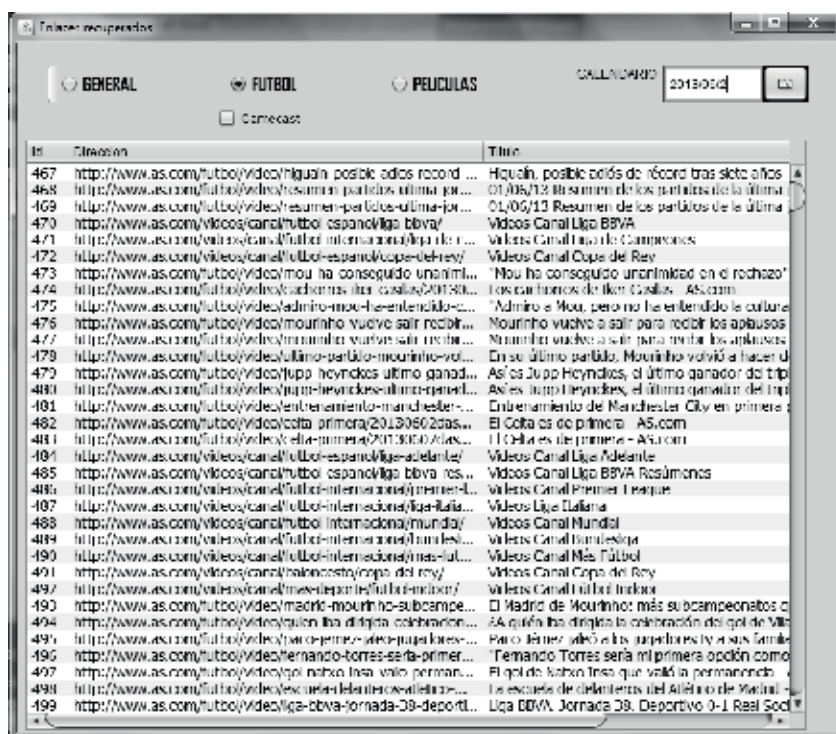


Fig. 4. Interfaz de enlaces recuperados.

- *Envío de datos al formulario.* Asigna y envía los valores semilla al formulario de búsqueda seleccionado para realizar la consulta. Los valores semilla son asignados al elemento HTML correspondiente del formulario analizado. Posteriormente, se realiza la simulación del evento clic en el elemento button mediante HtmlUnit y Selenium.
- *Analizador de respuesta.* Evalúa la respuesta obtenida a partir de la consulta realizada al formulario de búsqueda a través de un análisis completo del contenido HTML de la página resultante. Cabe mencionar que la mayoría de los buscadores presentan los resultados de búsqueda en un mecanismo de paginación; para este tipo de buscadores, solamente se analiza el contenido de la primera página de resultados y tomando en consideración los resultados de mayor relevancia encontrados en la página. La respuesta apropiada se almacena en la BD.

El resultado final del proceso realizado por esta arquitectura es reflejado en los registros de una tabla, la cual contiene un conjunto de enlaces con información relacionada con las palabras semilla.

#### 4. IMPLEMENTACIÓN

La Figura 3 muestra la interfaz principal de CrawNet desarrollado bajo la plataforma Java. La interfaz gráfica permite al usuario recuperar información de manera asistida. Los enlaces resultantes pueden ser visualizados sin necesidad de concluir el proceso de búsqueda de información.

Como puede apreciarse en la interfaz, la semilla URL es agregada en un campo de texto; el usuario podrá seleccionar diferentes tipos de búsqueda, tales como general, películas, videos, gamecast y películas avanzadas. Después que el usuario seleccione sus opciones de búsqueda, el usuario podrá ejecutar la consulta.

El resultado de la búsqueda es una lista de enlaces recuperados de una consulta en un dominio específico. En la Figura 4 puede apreciarse un ejemplo de los resultados de enlaces recuperados en una búsqueda de videos de fútbol.



Fig. 5. Implementación de búsqueda avanzada.

La implementación de la herramienta CrawNet fue realizada en sitios reales de internet. En el campo de semilla URL se agrega una dirección URL relacionado con cualquiera de cuatro tipos de búsqueda y, en cualquiera de los casos, se retorna una lista de enlaces recuperados (Figura 4). Para el caso de búsqueda en la web oculta, se realiza a través de la búsqueda avanzada donde se requiere rellenar campos para la búsqueda (Figura 5).

#### 5. EVALUACIÓN DE RESULTADOS

Para la evaluación del crawler se aplicaron las métricas de precisión, exhaustividad y medida F [22]. Estas métricas son utilizadas en la evaluación de los sistemas de recuperación de información (SRI). La precisión determina el porcentaje de aciertos de una operación en un SRI [23]. La exhaustividad determina el porcentaje de datos relevantes recuperados respecto al total de datos relevantes existentes en una BD [23]. La medida F (o medida armónica f) determina el promedio existente entre los valores de precisión y la exhaustividad [22].

Como se ha mencionado anteriormente, el algoritmo de la web superficial permite búsquedas generales, videos de fútbol, información de películas y gamecast. En el caso de búsquedas generales, no se aplicaron la precisión y exhaustividad por la carencia de filtros; sin embargo, se analizó el tiempo de consulta y el número de enlaces recuperados dentro y fuera del dominio (ver Tabla 2). En la Tabla 2 solo se muestra el resultado de los primeros 5 de los 120 enlaces evaluados.

**Tabla 2.** Resultados de caso de estudio de búsqueda general.

Sitio	Número de enlaces recuperados			
	En el dominio		Fuera del dominio	
	1 hora	2 horas	1 hora	2 horas
http://msn.foxsports.com	8178	10622	9033	18170
www.sport.es	2633	3380	8985	11481
www.cnn.com	2634	4297	12720	18720
www.amazon.com	10646	17483	22375	30837
www.youtube.com	18089	21443	14240	24170

En la Tabla 3 se puede apreciar los dominios seleccionados para los otros tipos de búsqueda. Se seleccionaron grupos de 100 enlaces que se utilizaron como muestra representativa de cada dominio; cada enlace fue almacenado en la BD para posteriormente realizar el análisis e Identificación de Elementos Relevantes (IER) y aplicar la métrica de precisión. Asimismo, la tabla muestra la IER Realizada Manualmente (IERM) y la IER mediante CrawNet (IERC) sobre las 100 muestras; por último, se aprecia el porcentaje de la Precisión (P) de la búsqueda.

La evaluación de los casos de estudio de la web oculta se dividió en “selección de formularios y elementos” (que evalúa la correcta selección del formulario) y “evaluación de respuesta generada” (que evalúa las respuestas generadas de las consultas realizadas por CrawNet). La evaluación de las dos secciones implicó una verificación directa sobre los dominios web con el fin de identificar formularios de búsqueda avanzada, para así registrar los URLs, etiquetas, elementos, etc. Posteriormente, se realiza la consulta para verificar los URLs que se obtenían como resultado.

### 5.1. Selección de formularios y elementos

Las pruebas se realizaron en 18 sitios (ver columna 1 de Tabla 4). La Tabla 4 también muestra un resumen de los resultados obtenidos de los 18 sitios en la identificación y selección de los elementos HTML (selección y recuperación del formulario y asociación de parámetros con valor de atributos en textbox y combobox). Puede apreciarse que 16 sitios obtuvieron el 100% en Precisión (P), Exhaustividad (E) y Medida F (M); en 2 sitios no se pudieron identificar formularios de búsqueda avanzada.

Por cada sitio, se realizaron dos tipos de verificaciones, manualmente y utilizando CrawNet. En la Tabla 5 se muestran los elementos descriptivos que sirvie-

**Tabla 3.** Resultados de la búsqueda en dominios específicos. Identificación de Enlaces Relevantes Manualmente (IERM), Identificación de Enlaces Relevantes mediante CrawNet (IERC).

Tipos de búsqueda	Dominio URL	IERM	IERC	P
Videos de fútbol	www.as.com	96	96/100	96%
	www.fifa.com	92	92/100	92%
	http://espnfc.com	95	95/100	95%
	http://futbol.univision.com	67	67/100	67%
	www.marca.com	100	100/100	100%
	Promedio total: 90%			
Info. de películas	www.accionhd.com	94	94/100	94%
	www.allmovie.com	100	100/100	100%
	www.amazon.es	88	88/100	88%
	www.elmultinice.com	100	100/100	100%
	www.estrenosdecine.net	100	100/100	100%
Promedio total: 96%				
Gamecast	http://espn deportes.espn.go.com	100	100/100	100%
	http://espnfc.com	100	100/100	100%
Promedio total: 100%				

ron para realizar la comparación entre los dos tipos de evaluaciones (columna 1). Debido a la falta de espacio, solo se muestra el análisis e identificación de parámetros del sitio www.filmsandtv.com; en dicho sitio, se seleccionó 1 de 2 formularios existentes en la página por lo que su P, E y M fue del 100%.

Los resultados obtenidos mediante la evaluación de las métricas de precisión y exhaustividad sobre la información recuperada arrojaron porcentajes muy altos, que, en la mayoría de los casos, fue de un promedio del 100%; sin embargo, este porcentaje se obtuvo sobre aquellos URLs que el algoritmo de la web oculta logró recuperar al identificar un formulario de búsqueda avanzada en su contenido. Los enlaces en los que no se llegó a identificar este tipo de formulario no pudieron ser evaluados. La razón por la que no se identificaron algunos formularios de búsqueda avanzada fue que existieron enlaces que no coincidieron con ningún patrón de búsqueda avanzada establecido.

**Tabla 4.** Resultados de la identificación y selección de elementos HTML. Selección y Recuperación del Formulario (SRF), Asociación de parámetros con valor de atributos en TextBox y Combobox (AP-TC) e Identificación y Selección de Botón (ISB).

Enlace URL	SRF			AP-TC			ISB		
	P	E	M	P	E	M	P	E	M
www.filmsandtv.com	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.findanyfil.com	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.half.ebay.com	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.starscfe.com	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.videomania.info/es	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.bbfc.co.uk	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.accionhd.com	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.swapadvd.com	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.chapter.indigo.ca	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.dvdgo.com	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.cinegratis.net	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.k12digitalmovies.com	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.movlic.com	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.megamovieline.com	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.movietickets.com	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.hispashare.com	100%	100%	100%	100%	100%	100%	100%	100%	100%
www.amazon.es	0%	0%	0%	0%	0%	0%	0%	0%	0%
www.amazon.com	0%	0%	0%	0%	0%	0%	0%	0%	0%
Total	88.8%	88.8%	88.8%	88.8%	88.8%	88.8%	88.8%	88.8%	88.8%

**Tabla 5.** Análisis e identificación de parámetros en formularios web del sitio www.filmsandtv.com.

Parámetros evaluados	Valores encontrados	
	Verificación manual	Verificación CrawNet
Enlace semilla	www.filmsandtv.com	www.filmsandtv.com
Enlace formulario	www.filmsandtv.com/advsearch.php	www.filmsandtv.com/advsearch.php
Formulario seleccionado	Name="frmAdvSrch"	Name="frmAdvSrch"
Número de etiquetas descriptivas relevantes	4	4
Dominio de etiquetas descriptivas relevantes	Title, actor, director, genre	Title, actor, director, genre
Número de elementos asignables relevantes	4	4
Dominio de elementos asignables	Título = "asTitle" Actor = "asActor1" Director = "asDirector1" Género = "asGenre"	Título = "asTitle" Actor = "asActor1" Director = "asDirector1" Género = "asGenre"
Botón seleccionado	Name="btnAdvSearch"	Name="btnAdvSearch"

**Tabla 6.** Resultados de evaluación de una película en 8 sitios web distintos con 3 Tipos de Búsqueda (TB): 1) por Título (T), 2) por Título y Actor (TA) y 3) por Título, Actor y Director (TD).

Enlaces URL	TB	TER	TERu	TERr	P	E	M
www.filmsandtv.com	T	6	101	6	5.9%	100%	11.14%
	TA	4	51	4	7.8%	100%	14.47%
	TD	2	31	2	6.4%	100%	12.03%
www.findanyfilm.com	T	15	23	15	65.22%	100%	78.95%
	TA	5	5	5	100%	100%	100%
	TD	3	3	3	100%	100%	100%
www.half.ebay.com	T	18	20	18	90%	100%	94.74%
	TA	18	12	10	83.33%	55.5%	66.67%
	TD	14	8	8	100%	100%	100%
www.bbfc.com.uk	T	20	81	15	18.52%	75%	30.07%
	TA	29	29	29	100%	100%	100%
	TD	9	9	9	100%	100%	100%
www.accionhd.com	T	7	8	7	87.50%	100%	93.33%
	TA	3	3	3	100%	100%	100%
	TD	0	0	0	-	-	-
www.swapadvd.com	T	10	26	3	11.54%	30%	16.67%
	TA	10	21	6	28.57%	60%	38.71%
	TD	5	6	5	83.33%	100%	69.77%
www.videomania.info	T	2	2	2	100%	100%	100%
	TA	1	2	1	50%	100%	66.67%
	TD	0	0	0	-	-	-
www.starscafe.com	T	10	174	10	5.75%	100%	10.87%
	TA	10	166	10	6.02%	100%	11.36%
	TD	10	3	2	66.67%	20%	30.77%

## 5.2. Evaluación de respuesta generada

Las pruebas realizadas en esta sección se realizaron en 8 sitios web (ver columna 1 de Tabla 6), los cuales fueron procesados por CrawNet. Por cada sitio se realizaron consultas de diez películas cinematográficas de forma manual y usando el algoritmo de la web oculta; sin embargo, debido a la carencia de espacio, únicamente se muestra el resultado de la búsqueda de una película (Título: Superman, Actor: Christopher Reeve, Director: Richard Lester) en cada sitio (Tabla 6).

En la Tabla 5 se muestra el resultado de la consulta de dicha película en la web oculta. Las pruebas en cada sitio web involucraron 3 tipos de búsqueda: búsqueda por título, búsqueda por título y actor y búsqueda por título, actor y director. Como primer paso, se analizó manualmente (en cada sitio) el Total de Enlaces Relevantes (TER), los cuales corresponden a los enlaces de las películas en cada sitio. Asimismo se evaluó el Total de Enlaces Recuperados

(TERu) por CrawNet y el Total de Enlaces Relevantes Recuperados (TERr) por CrawNet. Finalmente, se muestran los resultados P, E y F.

Los resultados obtenidos basados en la información recuperada arrojaron porcentajes muy altos en algunos casos. Los resultados sobre los ocho dominios web procesados siguieron un mismo comportamiento. CrawNet recuperó muchos más enlaces que los que debía recuperar, afectando con esto la precisión; sin embargo, se puede afirmar que se lograron recuperar todos los URLs esperados, dando como consecuencia porcentajes muy altos en cuanto a la exhaustividad del algoritmo.

## 6. CONCLUSIONES Y TRABAJOS FUTUROS

En las búsquedas permitidas en el algoritmo de la web superficial, se utilizaron patrones de expresiones regulares para poder filtrar la información recuperada. Se analizaron un total de 1200 URLs de forma manual para corroborar que la recuperación



fuera de acuerdo a la búsqueda seleccionada. Los resultados obtenidos de la web superficial fueron superiores al 90% en precisión; por lo tanto, se comprobó la efectividad del algoritmo desarrollado.

El algoritmo de la web oculta presentó retos durante su desarrollo ya que debía identificar páginas que tuvieran formularios de búsqueda avanzada, determinar el dominio de películas al que pertenecía y comprobar que el algoritmo lograra enviar datos al formulario para analizar la respuesta generada a dicha consulta.

El algoritmo de la web oculta logró identificar formularios de búsqueda avanzada en la mayoría de los casos; sin embargo, los resultados presentados fueron a partir de solo 18 sitios que permitían la búsqueda avanzada de películas cinematográficas. En los resultados se pueden apreciar que tuvo un 100% en las métricas de exhaustividad. Para estas pruebas puede parecer bastante alto, pero es importante aclarar que el porcentaje reportado es solo de los sitios en los que se pudo identificar adecuadamente el formulario de búsqueda avanzada; los demás sitios web en los que no se pudo identificar, no se tomaron en cuenta.

Actualmente se está trabajando en la generación de formularios web de manera automática, utilizando técnicas y metodologías actuales; asimismo, pretendemos utilizar ontologías para analizar, estructurar y organizar información de tal manera que permita descubrir patrones que siguen ciertos temas en particular para lograr la generación virtual de estos formularios y, una vez generados, utilizar CrawNet para realizar búsquedas de un tema en particular utilizando como plantillas de búsqueda los formularios generados.

## REFERENCIAS

- [1] H. Yeye, X. Dong, G. Venkatesh, R. Sriram & S. Nirav, "Crawling Deep Web Entity Pages", Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. Rome, Italy, pp. 355-364, 2013
- [2] M. Bergman, "White Paper: The Deep Web Surfacing Hidden Value". BrightPlanet: The Journal of Electronic Publishing, vol. 7, no. 1, 2012.
- [3] M. Álvarez. "Arquitectura para Crawling dirigida de información contenida en la web oculta". PhD. Dissertation, *Universidad Coruña*, A. Coruña, España, 2007.
- [4] S. Lawrence & C. Giles, "Accessibility of Information on the Web", *Nature*, vol. 400, no. 1, pp. 107-109, Julio, 1999.
- [5] B. Fernández. & S. Pardo, "Selección de recursos de información disponibles en el Web invisible". *Acimed*. vol. 14, no. 6, 2006.
- [6] Z. Wu, L. Jiang, Q. Zheng & J. Liu, "Learning to surface deep web content", In Proc. 2011, Twenty-Fourth AAAI Conference on Artificial Intelligence. *Georgia, USA*, pp. 1967-1968.
- [7] W. Yan, "Query selection in deep web crawling: Help your crawler efficiently retrieve data from the largest data sources in the web year", Scholar's Press, 2014.
- [8] K. Chang, B. He, & Z. Zhang, "Toward large scale integration: Building a MetaQuerier over databases on the web", 2005. Proceedings of the Second Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, January, pp. 44-55.
- [9] B. He, M. Patel, Z. Zhang & K. Chen-Chuan, "Accessing the deep web: A survey", *Commun. ACM*, vol. 50 no. 5, pp 94-101. Mayo, 2007.
- [10] M. Soulemane, M. Rafiuzzaman & H. Mahmud, "Crawling the hidden web: An approach to dynamic web indexing", *International Journal of Computer Applications*, vol. 55, no. 1, pp 7-15, Octubre, 2012.
- [11] D. Anuradha & A. Babita, "Hidden web extractor: Dynamic way to uncover the deep web", *International Journal on Computer Science & Engineering*, vol. 4, no. 6, pp. 1137-1145. Junio, 2012.
- [12] L. Xian, D. Xin, L. Kenneth, M. Weiyi & S. Divesh, "Truth finding on the deep web: Is the problem solved?", Proceedings of the 39th international Conference on Very Large Data Bases, Trento, Italy, pp. 97-108, 2013

- [13] S. Liddle, S. Yau & D. Embley, "On the automatic extraction of data from the hidden web". In Proc. 2001 *International Workshop on Data Semantics in Web Information Systems (DASWIS-2001)*, London, UK, pp. 212-226.
- [14] C. Sherman & G. Price. 2001, "The invisible web: Uncovering information sources search engines can't see". Medford, N.J, CyberAge Books, 2001.
- [15] M. Álvarez, J. Raposo, A. Pan, F. Cacheda, F. Bellas, & V. Carneiro. "Crawling the content hidden behind web forms", Proceedings of the ICCSA, Lecture Notes in Computer Science v 4706, Springer, pp. 322-333, 2007
- [16] V. Prieto, M. Álvarez, R. López-García & F. Cacheda, "A scale for crawler effectiveness on the client-side hidden web", *Computer Science and Information Systems*, vol. 9 no. 2, pp. 561-583. Junio, 2012.
- [17] D. Lewandowski & P. Mayr, "Exploring the academic invisible web". *Library Hi Tech.*, vol. 24, no. 4, pp. 529-539, Feb. 2007.
- [18] M. Wuand & A. Marian, "A framework for corroborating answers from multiple web sources". *Information Systems*, vol. 36, no. 2, pp. 431-449, Jun. 2011.
- [19] X. Dong, B. Saha & D. Srivastava, "Less is more: Selecting sources wisely for integration". *PVLDB*, vol. 6, no. 2, 2013. Disponible en: <http://www.vldb.org/pvldb/vol6/p37-dong.pdf>
- [20] S. Raghavan, & H. Garcia-Molina, "Crawling the hidden web," Proceedings of the 27th International Conference on Very Large Data Bases (VLDB 2001), San Francisco, CA, USA, pp. 129-138, 2001
- [21] M. Cafarella, E. Chang, A. Fikes, A. Halevy, W. Hsieh, A. Lerner, J. Madhavan & S. Muthukrishnan, "Data management projects at Google". *ACM SIGMOD Record* vol. 37, no. 1, pp.34-38, 2008.
- [22] Salinas Martínez, Osvaldo. "Modelado semántico de documentos con estructura definida". Tesis, Cd. Victoria, Tamaulipas, México, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, 2012.
- [23] F. Martínez-Méndez. *Recuperación de información: Modelos, sistemas y evaluación*, Murcia, España, Ed. El Kiosko, 2012.



# Paralelización del Algoritmo Expectación-Maximización Utilizando OpenCL

## Maximization of Expectation-Parallelism Algorithm Using OpenCL

**Ernesto Insua-Suárez, Ing.**

*Facultad de Ingeniería Informática  
Instituto Superior Politécnico José Antonio Echeverría  
La Habana, Cuba  
einsuas@ceis.cujae.edu.cu*

**Marlis Fulgueira-Camilo, Ing.**

*Facultad de Ingeniería Informática  
Instituto Superior Politécnico José Antonio Echeverría  
La Habana, Cuba  
mfulgueira@udio.cujae.edu.cu*

**Venus Henry-Fuenteseca, Ing.**

*Facultad de Ingeniería Informática  
Instituto Superior Politécnico José Antonio Echeverría  
La Habana, Cuba  
vhenry@ceis.cujae.edu.cu*

(Recibido el 10-09-2014. Aprobado el 11-12-2014)

**Resumen.** Actualmente, las organizaciones y empresas almacenan grandes volúmenes de datos para lograr sus propósitos. Una de las variantes para obtener información valiosa consiste en el empleo de la minería de datos. Dentro de esta, existen diferentes tareas, una de ellas es el agrupamiento. En esta tarea, los datos se agrupan según sus semejanzas entre sí y sus diferencias con elementos de otros grupos. Dentro de los algoritmos que realizan estos agrupamientos se encuentra expectación-maximización, el cual presenta elevados tiempos de ejecución en la medida que aumenta el tamaño de los datos. En el presente artículo, se discute acerca de la paralelización del algoritmo, utilizando técnicas de programación paralela. El diseño del algoritmo propuesto se basa en el uso de las tarjetas de procesamiento gráfico, GPU. OpenCL, lenguaje empleado para la programación en arquitecturas híbridas, permite aprovechar las arquitecturas de hardware dis-

ponibles, con lo que se logra disminuir el tiempo de ejecución de la implementación realizada. La razón principal por la cual es posible mejorar este tiempo se debe a la cantidad de procesos paralelos que se pueden lanzar en hilos de procesamientos independientes. Para el logro de los resultados descritos, se integran conocimientos del campo de la minería de datos y la computación paralela y distribuida. Como parte de esta investigación, se realizó una implementación del algoritmo, utilizando las bibliotecas de OpenCL, para disminuir su tiempo de ejecución. La implementación logra disminuir en un 82% la implementación secuencial. Esto significa que el algoritmo paralelo se ejecuta 5,5 veces más rápido que su correspondiente implementación secuencial.

**Palabras claves:** Arquitecturas híbridas, computación paralela y distribuida, expectación-maximización, ejecución paralela, OpenCL.

**Citación de artículo, estilo IEEE:**

E. Insua-Suárez, M. Fulgueira-Camilo, V. Henry-Fuenteseca, "Paralelización del Algoritmo Expectación-Maximización Utilizando OpenCL", *Lámpsakos*, N° 13, pp. 51-61, 2015.

DOI: <http://dx.doi.org/10.21501/21454086.1361>

**Abstract.** Nowadays both organizations and companies store big volumes of data to achieve their purposes. One of the variants to obtain valuable information consists on the employment of Data Mining. Inside Data Mining, different tasks exist and one of them is clustering. In this task the data group according to their likenesses among them differ with elements of other groups. One of the algorithms that carry out these clusters is Expectation-Maximization, which presents high times of execution in their data. This article discusses about the parallelization of the mentioned algorithm, using techniques of parallel programming. The design of the proposed algorithm is based on the use of the graphic process unit, GPU. OpenCL, language used for the programming in hybrid architectures, allows to take advantage of the available hardware architectures, which it is possible to diminish the time of execution of the sequen-

tial implementation. The reason to improve this time is due to the quantity of parallel processes that can rush in threads of independent prosecutions. For the achievement of the described results, knowledge of the field of Data Mining and Parallel and Distributed Computation are integrated. As part of this investigation, an implementation of the algorithm using the libraries of OpenCL was carried out to diminish the time of execution. The implementation is able to diminish the sequential implementation in 82%, this means that the parallel algorithm is executed 5,5 times quicker that its sequential corresponding implementation.

**Keywords:** Hybrid architectures, Parallel & Distributed Computation, Maximization of Expectation, Parallel Execution, OpenCL.

## 1. INTRODUCCIÓN

Con el transcurso del tiempo, las organizaciones han necesitado obtener información procedente de los datos que manejan. Con el creciente volumen de estos datos, procesarlos se convierte en un trabajo complicado. Una de las variantes existentes para realizar este procesamiento es la minería de datos, la cual se puede definir como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos [1].

La minería de datos está compuesta por varios tipos de tareas, una de estas es el agrupamiento. El agrupamiento consiste en obtener grupos a partir de los datos, los cuales se basan en el principio de que la similitud entre los elementos de un grupo debe ser la mayor posible y que la semejanza entre elementos de grupos diferentes debe ser mínima [1].

La presente investigación, se enfoca en la reducción del tiempo de ejecución de la implementación del algoritmo de minería de datos expectación-maximización (EM). Este algoritmo de agrupamiento fue presentado por primera vez por Laird y Rubin Dempster en la revista *J Royal Statistical Society*, en 1977 [2]. Expectación-maximización selecciona el número de grupos a obtener. Se utiliza como parte del preprocesamiento de los datos, para luego aplicar el algoritmo de agrupamiento K-MEANS con la cantidad de grupos obtenida con EM. Dado que el algoritmo es parte de una etapa de preprocesado, se necesita que se ejecute de manera rápida para luego realizar el agrupamiento. Además, con la creciente cantidad de datos a procesar, el factor tiempo viene a jugar un papel decisivo.

Dadas las condiciones anteriores, una de las mejores alternativas de solución la podemos encontrar en la explotación de las características de procesamiento simultáneo que presentan los diferentes dispositivos multinúcleos actuales, los cuales pueden ser procesadores o tarjetas de procesamiento gráfico. Para utilizar la característica de procesamiento simultáneo de estos dispositivos es necesario utilizar una biblioteca que permita la distribución de las tareas entre los diferentes núcleos. La presente investigación muestra los resultados obtenidos luego de utilizar una de estas bibliotecas de procesamiento paralelo en comparación con una implementación secuencial del algoritmo EM, para lo cual se realizaron las implementaciones secuenciales y paralelas del algorit-

mo. Con la aplicación del procesamiento paralelo, el principal beneficio que se logra es la disminución del tiempo de ejecución de los algoritmos.

La problemática a resolver será el procesamiento de un fichero .xes al cual se le desean realizar agrupamientos de sus datos.

Esta investigación representa una segunda etapa de implementación del algoritmo Expectación-Maximización, la primera etapa estuvo constituida por la implementación del algoritmo utilizando las bibliotecas de OpenMP. La actual etapa tiene como principal propósito introducir una nueva biblioteca que permita mejorar el tiempo de ejecución del algoritmo antes mencionado.

## 2. DESARROLLO DEL ARTÍCULO

### 2.1. Minería de datos

Dentro de la minería de datos, se pueden distinguir varios tipos de tareas, las cuales se emplean en dependencia de cómo se desea obtener el nuevo conocimiento o qué procesamiento se le desea realizar a los datos. Estas tareas pueden ser de tipo descriptivas o predictivas: las descriptivas sirven para conocer el comportamiento de los datos y las predictivas se utilizan para estimar valores futuros de las variables. Dentro de las principales tareas se pueden encontrar:

#### A. Predictivas

- Regresión: Consiste en asignarle a una variable un valor real. Solo se utiliza para realizar predicciones de variables numéricas.
- Clasificación: Se clasifica el elemento que se esté analizando, utilizando los atributos propios del elemento. Su principal diferencia con la regresión consiste en que la estimación se puede realizar sobre cualquier tipo de variable.

#### B. Descriptivas

- Correlación: Consiste en realizar un análisis con el fin de obtener la magnitud de la similitud de una variable con otra.

- Reglas de asociación: Tiene cierta similitud con la correlación, pero su principal objetivo es identificar relaciones no explícitas entre varios atributos: por ejemplo, “Si nombre es igual Ana, entonces sexo es femenino”.
- Agrupamiento: También conocido como *clustering*. Consiste en el proceso de obtener grupos a partir de un conjunto de datos, maximizando la similitud entre los elementos de un grupo y haciéndola mínima con respecto a los elementos de otros grupos [1].

El algoritmo expectación-maximización es un algoritmo de minería de datos que se encuentra dentro de la tarea de agrupamiento.

El agrupamiento es el proceso de examinar una colección de elementos y agruparlos en grupos o “clusters”, de acuerdo a alguna medida de distancia. El objetivo a lograr en las tareas de agrupamiento es que los elementos dentro de un grupo tengan la menor distancia o diferencia posible uno de otro, como se observa en la Figura 1.

## 2.2. Expectación-Maximización

El algoritmo de expectación-maximización (EM) se ha convertido en una herramienta popular para realizar agrupamientos en grandes volúmenes de datos. Existen dos aplicaciones principales del algoritmo EM. La primera es cuando los datos tienen valores faltantes derivados del proceso de observación y la segunda está dada por la posibilidad de la estimación de patrones. Este algoritmo fue presentado por primera vez por Laird y Rubin Dempster en J Royal Statistical Society en 1977.

Debido a su implementación, el algoritmo EM puede ser visto como un método metaheurístico, garantizando que se encuentre, con cierta probabilidad, un máximo local del registro de los datos. EM brinda una aproximación estadística al problema buscando el número de “cluster” más probable dados los datos. La base de este tipo de agrupamiento se encuentra en un modelo estadístico llamado mezcla de distribuciones, donde cada distribución representa la probabilidad de que un objeto tenga un conjunto particular de pares atributo-valor. El algoritmo EM es un procedimiento iterativo que calcula la máxima probabilidad (MP) estimada en presencia de datos faltantes u ocultos.

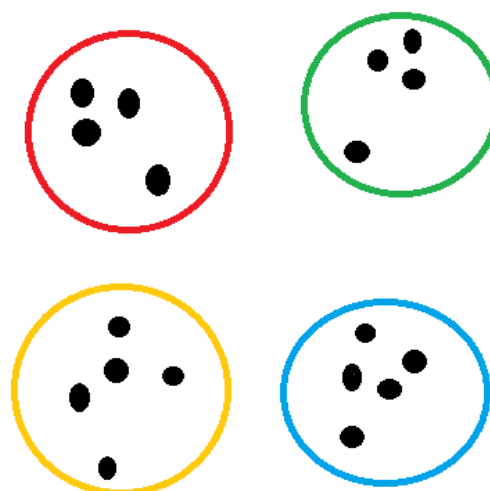


Fig. 1. Agrupamiento de elementos.

Cada iteración del algoritmo EM consiste en dos procesos: el E-paso y el M-paso.

En la expectación, o el E-paso, los datos faltantes son calculados con los datos observados y la estimación actual del modelo de los parámetros. Este archivo se realiza usando la expectativa condicional.

En el M-paso, la función de probabilidad es maximizada bajo la suposición de que los datos faltantes son conocidos. La estimación de los datos faltantes del E-paso se usa en lugar de los datos faltantes reales.

El algoritmo detiene las iteraciones cuando ocurre la convergencia, lográndose esta cuando al calcular la verosimilitud (relación entre elementos), esta sea menor que un valor predefinido [1-3].

### 2.2.1. Idea general del algoritmo en pseudocódigo

En la Figura 2, se muestra el pseudocódigo del algoritmo expectación máxima.

### 2.2.2. Campos de aplicación

EM ha sido exitosamente aplicado en problemas de aprendizaje de máquina y reconocimiento de patrones en general. Generalmente, se utiliza como algoritmo de clustering (agrupamiento), en donde los puntos a clasificar se asumen como si estos fueran

**ALGORITMO EM ( $B=(G, \theta)$ ): Red Bayesiana,  $D$ : Datos)****Fase de Inicialización:**

Inicializar el conjunto de parámetros

Inicializar contador de etapas  $e=0$ **Fase Iterativa:****MIENTRAS** no convergencia

Paso E: Etapa de cálculo de Esperanzas

Paso M: Etapa de maximización

**FIN MIENTRAS****FIN ALGORITMO**

Fig. 2. Seudocódigo del algoritmo expectación-maximización.

derivados de una mezcla de componentes gaussianas o Poisson y el interés es encontrar los parámetros de dicha distribución [1-3].

**2.2.3. Escenarios de utilización**

- Visión computacional (para etiquetar regiones de imágenes: tipos de tejidos).
- Biología computacional (inferencia de árboles filogenéticos).
- Minería de datos (análisis de bases de datos).
- Finanzas (manejo de riesgos y composición de portafolios).

**2.3. Variantes**

Varios métodos se han propuesto para acelerar la lenta convergencia del algoritmo EM, como los que utilizan el declive conjugado y modificaron técnicas de Newton-Raphson.

*La expectativa maximización condicional* (ECM) sustituye cada M de paso por una secuencia de pasos de maximización condicional (CM) que cada parámetro  $\theta$  se maximiza individualmente, con reservas en los otros parámetros restantes fijado [4].

*Maximización de la expectativa generalizada* (GEM), en la cual uno solo busca un aumento de la función objetivo  $F$  tanto para el paso de E como para M, según la descripción alternativa [4].

**2.4. Solución propuesta para resolver el problema**

Una de las alternativas existentes es hacer uso del paralelismo, el cual consiste en la ejecución de manera simultánea, en un instante de tiempo, de sentencias de un lenguaje de programación específico, realizando un mejor aprovechamiento de las arquitecturas que actualmente existen.

Para lograr un nivel de paralelismo lo más eficiente posible, se siguió la metodología de diseño de algoritmos paralelos enunciada por Ian Foster [5]. Esta está constituida por cuatro pasos a seguir en el momento que se diseña un algoritmo para ejecutar de forma paralela. Los cuatro pasos son [5]:

- **Particionado:** Se dividen los datos o las funciones en dependencia de las características del algoritmo a paralelizar sin tener en cuenta la arquitectura sobre la cual se realizará la ejecución.
- **Comunicación:** Se realiza un análisis referente a cómo se realizan las comunicaciones entre los datos o funciones en la implementación secuencial.
- **Agrupamiento:** Se agrupan los datos ya estando particionados, teniendo en cuenta la arquitectura del sistema de cómputo y las características de comunicación analizadas en el paso anterior.
- **Mapeo:** Se le asignan a cada hilo de ejecución de la arquitectura seleccionada, cada grupo de datos o funciones obtenidos en el paso anterior.

Para lograr cumplir el objetivo de disminuir el tiempo de ejecución del algoritmo EM, se deben seguir una serie de pasos adicionales luego de la implementación secuencial del algoritmo:

- Encontrar las secciones del código que ocupan mayor tiempo al procesador; para esto se utilizan herramientas como Intel Parallel Studio.
- Con las secciones seleccionadas en el paso anterior, realizar un análisis con el fin de descubrir si estas secciones permiten ser paralelizadas. Secciones paralelizables son aquellas

en la que unos datos no dependan de otros; si existiera dependencia, no se podría aplicar paralelismo [5].

- Luego de tener filtradas estas secciones, se procede a realizar las transformaciones correspondientes al proceso de paralelismo que se vaya a utilizar.

Como resultado del seguimiento de los pasos definidos anteriormente, se decidió aplicarle el proceso de transformación al paso E, debido a que es el paso que más procesamiento necesita y el cual permite, además, que el paralelismo sea aplicado.

## 2.5. Alternativas de implementación paralela

Para cada una de las alternativas analizadas a continuación se necesita realizar un estudio de las transformaciones que es necesario aplicar al lenguaje para hacer uso del paralelismo y un mejor aprovechamiento de las arquitecturas de procesamiento disponibles en función de la alternativa de implementación que se desee utilizar.

**OpenMP** (Open Multi-Processing) es una biblioteca para implementaciones paralelas basada en el uso de memoria compartida [6, 7].

Requisitos: Se necesita poseer un microprocesador multinúcleo para poder ejecutar las sentencias de manera simultánea.

El paradigma de memoria compartida consiste en que múltiples núcleos de un dispositivo acceden a la misma memoria que, en este caso, es la memoria caché de los microprocesadores disponibles, al contrario del paradigma de memoria distribuida, donde se presentan varios núcleos, cada uno con su propia memoria local y que cooperan en conjunto para resolver tareas mediante alguna herramienta para administrar la ejecución de la tarea [8, 9].

**CUDA** (Dispositivo de Arquitectura Computacional Unificada o **Compute Unified Device Architecture**) es una biblioteca para implementaciones paralelas creada por la compañía NVIDIA y basada en el uso de GPUs (tarjetas de procesamiento gráfico) de dicha compañía [10, 11].

Requisitos: Se necesita poseer una tarjeta procesadora de gráficos de la marca NVIDIA.

**OpenCL** (Open Computing Language o lenguaje de computación abierto) consta de una interfaz de programación de aplicaciones y de un lenguaje de programación. OpenCL permite crear aplicaciones con paralelismo a nivel de datos y de tareas que pueden ejecutarse tanto en unidades centrales de procesamiento como en unidades de procesamiento gráfico [12-15].

Un kernel de OpenCL es una porción paralela (muy similar a una función) que es ejecutada en el device (tarjeta de procesamiento gráfico o unidades centrales de procesamiento). Este kernel puede ser ejecutado por muchos hilos en un instante de tiempo [12-14].

Un thread o hilo (en inglés) se ejecuta en el device mediante un arreglo de ellos que ejecutan el mismo código kernel, teniendo cada hilo un identificador (id) para su identificación y control. Un hilo es la unidad básica para el procesamiento paralelo. Los hilos pueden ser de dos tipos: físicos y virtuales [12-14].

Los hilos pueden efectuar comunicaciones entre sí en un momento dado, ya sea para compartir resultados o accesos a memoria, reduciendo así el ancho de banda. Estos hilos dentro de un bloque cooperan mediante la memoria compartida, pudiendo realizarse su sincronización en caso de que se requiera. Cada hilo tiene su propia memoria local. La Figura 3 muestra cómo se encuentra estructurado un device.

La memoria local está localizada en el dispositivo DRAM de la PC donde se ejecuta el algoritmo, por lo cual se debe realizar un acceso mínimo a esta memoria para reducir el tiempo de ejecución. Se debe siempre tratar que entre accesos a la memoria DRAM, se realicen la mayor cantidad de procesamientos posibles en el dispositivo, con el objetivo de minimizar el tiempo de comunicación [13, 15].

Se decide utilizar las bibliotecas de OpenCL para el procesamiento paralelo, pues se puede estar utilizando en una sola implementación CPU o GPU. OpenCL es libre de licencias, por lo cual una aplicación implementada utilizando esta biblioteca se pudiera estar comercializando sin ningún problema.

## 2.6. Arquitectura y fichero de prueba

El fichero que contiene los datos sobre los que se realizará el agrupamiento es de extensión .xes. Los ficheros .xes son utilizados en la minería de proce-



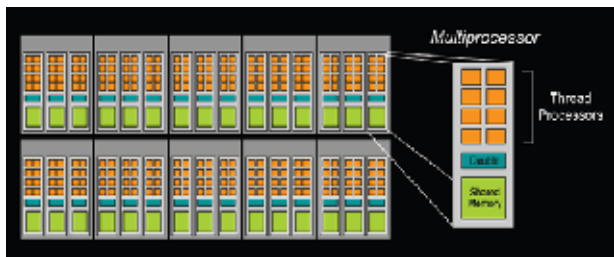


Fig.3. Estructura de un device.

Los archivos para registrar eventos de un proceso específico. La estructura de los ficheros .xes es muy parecida a la que presentan los de extensión .xml.

En este caso, el fichero .xes utilizado se corresponde con un proceso que ocurre dentro de un hospital, donde se desean agrupar los datos en base a valores numéricos.

El agrupamiento se realizará sobre la variable "Age", que contiene 2071 instancias en el fichero.

Las arquitecturas donde se probará el algoritmo es la siguiente:

Estas arquitecturas de prueba fueron seleccionadas con base en las diferencias en cuanto a la cantidad de núcleos que estas presentan. Además, se toman ya que son unas de las arquitecturas más comunes que se pueden encontrar en el mercado.

### 3. RESULTADOS

#### 3.1. OpenMP

Las figuras 4, 5 y 6 muestran las comparaciones de los resultados obtenidos con las ejecuciones paralela y secuencial, y sus respectivas representaciones gráficas del uso de cada CPU, ejecutándose el algoritmo sobre la tercera arquitectura.

Tabla 1. Arquitecturas empleadas.

Computadora	Núcleos/hilos	Memoria RAM	Frecuencia
Intel i7 860	4/8	8 GB	2.8 GHz
Intel Core 2 Quad Q9300	4/4	4 GB	2.5 GHz
Intel Core 2 Duo E7300	2/2	2 GB	2.66 GHz

Tabla 2. Características de las GPUs empleadas

GPU	Cantidad de núcleos
nVidia GeForce GTX 460	336
nVidia GeForce GTX 550 Ti	192
nVidia GeForce GTX 260	216



Fig. 4. Resultados de la ejecución secuencial.



Fig. 5. Resultado de la ejecución utilizando OpenMP.

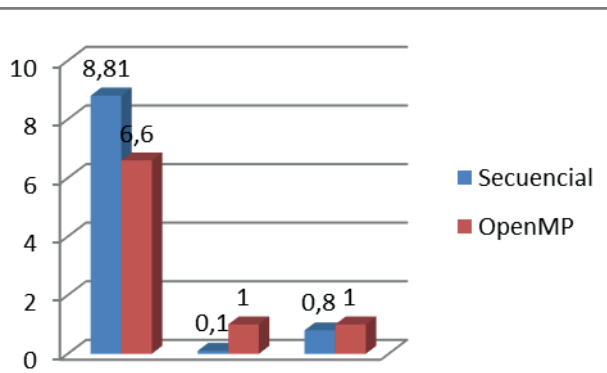


Fig. 6. Comparación de las ejecuciones secuencial y paralela utilizando OpenMP.

- Se obtuvo una disminución del 35% del tiempo de ejecución secuencial, lo cual significa que se está reduciendo considerablemente el tiempo de ejecución con la nueva alternativa de implementación.
- Se obtuvo el mismo patrón de los datos, lo que quiere decir que el algoritmo funciona correctamente con ambas implementaciones.
- Se realizó un aprovechamiento mayor de las capacidades multiprocesamiento del medio de cómputo ya que ambos núcleos fueron aprovechados al máximo de su capacidad y se redujo considerablemente el tiempo ocioso o de espera de los procesadores para la asignación de tareas.
- El algoritmo implementado de forma paralela se ejecuta 1,33 veces más rápido que el secuencial.

### 3.2. OpenCL

La Figura 8 muestra las comparaciones de los resultados obtenidos con las ejecuciones paralela y secuencial.

Nota: La ejecución del algoritmo utilizando OpenCL se realizó utilizando la GPU porque con este dispositivo se obtienen los mejores resultados, al tener muchas más unidades de procesamiento que un CPU.

```
Medias y desviaciones estandar por grupos
Grupo 0
Media 79.141661 Desviacion estandar 1.014044
Grupo 1
Media 83.850125 Desviacion estandar 2.255404
Grupo 2
Media 38.134040 Desviacion estandar 7.360293
Grupo 3
Media 70.836503 Desviacion estandar 4.656984
Grupo 4
Media 45.894234 Desviacion estandar 5.105981
Grupo 5
Media 59.758005 Desviacion estandar 3.533770
Grupo 6
Media 88.222409 Desviacion estandar 4.659782
Grupo 7
Media 59.758005 Desviacion estandar 3.533770
Grupo 8
Media 52.279891 Desviacion estandar 1.912773
Grupo 9
Media 45.894234 Desviacion estandar 5.105981
Plataforma: NUIDIA CUDA
Nombre dispositivo: GeForce GTX 550 Ti
Versimilitud -8557.219630
Duracion
valor 84.611000 segundos
```

Fig. 7. Resultado de la ejecución utilizando OpenCL.

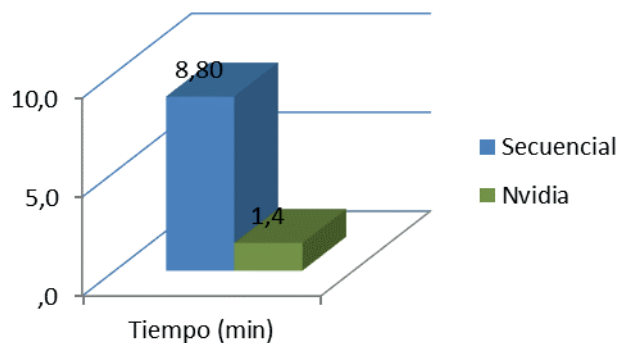


Fig. 8. Comparación de las ejecuciones secuencial y paralela utilizando OpenCL.

La Figura 9 y Figura 10 muestran los diferentes tiempos de ejecución utilizando OpenCL sobre CPU y GPU.

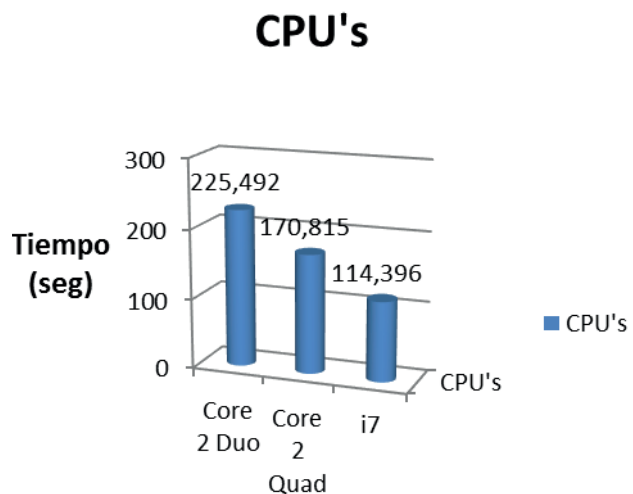


Fig. 9. Comparación entre CPUs OpenCL.

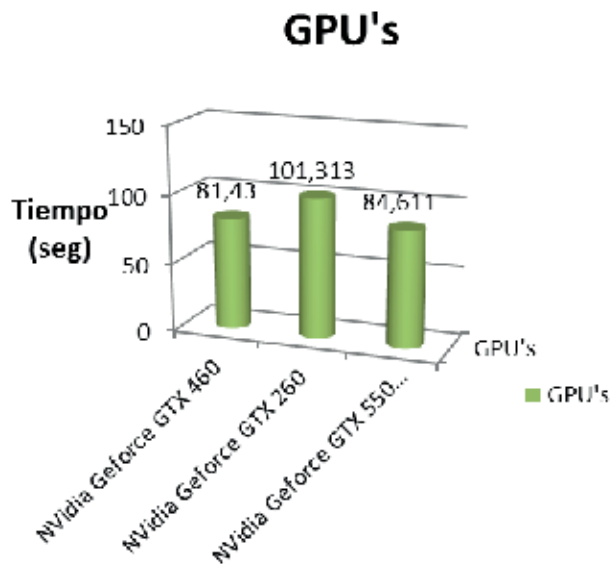


Fig. 10. Comparación entre GPUs OpenCL.

Se obtuvieron los mismos resultados en las 2 ejecuciones, o sea que la ejecución es correcta, partiendo de que la implementación secuencial también lo es.

El tiempo de ejecución de la plataforma NVIDIA fue de 1,4 minutos, que representa solo el 15% del tiempo obtenido por la ejecución secuencial; esto significa que se disminuyó el tiempo de ejecución en un 85%.

Se realizó un aprovechamiento mayor de las capacidades heterogéneas de cómputo disponibles.

El algoritmo implementado de forma paralela se ejecuta 6,3 veces más rápido que el secuencial.

La Tabla 3 muestra un resumen de los valores obtenidos con las diferentes bibliotecas empleadas.

#### 4. CONCLUSIONES

El aprovechamiento de las arquitecturas actuales permite que se realicen ejecuciones de algoritmos en menor tiempo; en especial, aquellos que trabajan con grandes volúmenes de datos, como es el caso de los algoritmos relacionados con la minería de datos. Para lograr un aprovechamiento de estas posibilidades, es necesario realizar estudio de la tecnología y del proceso al cual se le desee realizar la paralelización. Se deben buscar cuáles son las zonas críticas en procesamiento, para así dirigir los esfuerzos hacia esa sección del proceso.

Tabla 3. Resumen de tiempos

Ejecución	Tiempo(min)	% reducción
Secuencial	8,8	0
OpenMP	6,6	35
OpenCL	1,4	85

Como resultado de esta investigación, se cuenta con una implementación del algoritmo expectación-maximización, que disminuye su tiempo de ejecución secuencial y que puede ser usada en procesos que requieran agrupamiento de los datos.

El menor tiempo de ejecución de los CPUs utilizados, lo presenta el Intel Core i7, mientras que el mayor tiempo de ejecución lo presenta el Intel Core 2 Duo. La diferencia de tiempos se debe a la cantidad de hilos de ejecución y frecuencias de reloj que tiene cada uno de los procesadores. El CPU Intel Core i7 es la elección si se desea ejecutar el algoritmo sobre CPU.

El menor tiempo de ejecución de las GPUs utilizadas lo presenta la tarjeta de video nVidia GeForce GTX 460, mientras que el mayor tiempo de ejecución lo presenta la nVidia GeForce GTX 260. La diferencia de tiempos se debe a la cantidad de hilos de ejecución, anchos de banda, frecuencias de reloj y memoria que tiene cada una de las tarjetas gráficas. Debido a que el algoritmo no es 100% paralelizable, la parte secuencial se realiza en el CPU; y cuando se ejecuta sobre la GPU nVidia GeForce GTX 460, la parte secuencial se ejecuta en el CPU Intel Core i7, que es el CPU más rápido. Teniendo en cuenta los aspectos anteriores, la elección, si se desea ejecutar el algoritmo sobre GPU, es nVidia GeForce GTX 550 Ti, ya que la diferencia de tiempo de ejecución entre esta y la 460 es poco significativa y esta arquitectura cuenta con el CPU más lento de los empleados, que es el Intel Core 2 Duo.

Existe una considerable disminución del tiempo de ejecución con respecto a la implementación secuencial, siendo el tiempo paralelo 2 o más veces más rápido que el secuencial.

## 5. RECOMENDACIONES

Se recomienda realizar una implementación paralela del algoritmo, enfocada en un ambiente distribuido, para utilizar los recursos de más de una computadora y así realizar una mayor disminución de los tiempos de ejecución obtenidos como resultado de la investigación.

## 6. ANEXO: GLOSARIO

CUDA: Siglas de Compute Unified Device Architecture (Arquitectura de Dispositivos de Cómputo Unificado). Hace referencia tanto a un compilador como a un conjunto de herramientas de desarrollo creadas por NVidia que permiten a los programadores usar una variación del lenguaje de programación C para codificar algoritmos en GPU de NVidia [10, 11, 16].

OpenCL: Open Computing Language, en español, Lenguaje de Computación Abierto. Consta de una interfaz de programación de aplicaciones y de un lenguaje de programación. Juntos permiten crear aplicaciones con paralelismo a nivel de datos y de tareas que pueden ejecutarse tanto en unidades centrales de procesamiento como en unidades de procesamiento gráfico [12-14].

OpenMP: Interfaz de programación de aplicaciones (API) para la programación multiproceso de memoria compartida en múltiples plataformas. Permite añadir concurrencia a los programas escritos en C, C++ y Fortran sobre la base del modelo de ejecución fork-join [6].

Expectación-Maximización: Se usa en estadística para encontrar estimadores de máxima verosimilitud de parámetros en modelos probabilísticos que dependen de variables no observables [2-4, 17].

Minería de datos: Consiste en la extracción no trivial de información que reside de manera implícita en los datos. Dicha información era previamente desconocida y podrá resultar útil para algún proceso [1, 18, 19].

Arquitecturas multinúcleo: Es el diseño conceptual y la estructura operacional fundamental de un sistema de computadora cuya unidad de procesamiento está compuesta por varios núcleos con uno o varios hilos de ejecución [5, 20].

Kernel: Un kernel de OpenCL es una porción paralela (muy similar a una función) que es ejecutada en el device. Los hilos pueden ser de dos tipos: físicos y virtuales [12-14].

Thread: Hilo de ejecución que va a tener cada uno de los núcleos del procesador [12-14].

Device: Tarjeta de procesamiento gráfico [12-14].

## REFERENCIAS

- [1] J. Hernández-Orallo, J. Ramírez-Quintana, *Introducción a la minería de datos*, Primera edición, Madrid, Pearson Prentice Hall S.A., 2004, 680p.
- [2] A. P. Dempster, D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society*, pp. 39-45, 1977.
- [3] Collins, M, *The EM Algorithm*, p. 28, 1997. Disponible en: [http://faculty.washington.edu/fxia/courses/LING572/EM\\_collins97.pdf](http://faculty.washington.edu/fxia/courses/LING572/EM_collins97.pdf)
- [4] S. Purcell, *Maximum Likelihood Estimation (MLE)*, 2007, Disponible en: [http://statgen.iop.kcl.ac.uk/bgim/mle/sslike\\_1.html](http://statgen.iop.kcl.ac.uk/bgim/mle/sslike_1.html)
- [5] I. Foster, *Designing and building parallel programs*, Addison-Wesley, 1995. Disponible en: <http://www.mcs.anl.gov/~itf/dbpp>
- [6] OpenMP.org. *The OpenMP API Specification for parallel programming*, 2014, Disponible en: <http://www.openmp.org>
- [7] R. Pas, "An Overview of OpenMP 3.0", IWOMP 2009, TU Dresden (Alemania), Disponible en: [https://iwomp.zih.tu-dresden.de/downloads/2.Overview\\_OpenMP.pdf](https://iwomp.zih.tu-dresden.de/downloads/2.Overview_OpenMP.pdf)
- [8] H. Hoeger, *Introducción a la Computación Paralela*, Centro Nacional de Cálculo Científico Universidad de Los Andes, Mérida (Venezuela)–CeCaCULA, 2011, p. 48.
- [9] M. Tosini, *Introducción a las arquitecturas paralelas*, 2011, p. 51. Disponible en: <http://bit.ly/1H9pylQ>

- [10] NVIDIA, High Performance computing with CUDA, Users Group Conference, San Diego, CA (USA), 2009. Disponible en: [www.nvidia.com](http://www.nvidia.com)
- [11] J.A. Raya-Vaquera, E.X. Martín-Rull, *Aceleración con CUDA de procesos 3D*, 2009. Disponible en: <http://bit.ly/1Gg7mRu>
- [12] Wang, P., OpenCL Optimization, GPU Technology Conference, San José, CA (USA), 2009.
- [13] A. Munshi, T. G. Mattson, J. Fung, D. Ginsburg, *OpenCL Programming Guide*, Boston, MA (USA), Pearson Education, 2012.
- [14] A. Barak, E. Levy, A. Shiloh, A Package for OpenCL Based Heterogeneous Computing on Clusters with Many GPU Devices, IEEE International Conference on Cluster Computing Workshops and Posters, Heraklion, Isla de Creta, 2010.
- [15] B. R. Gaster, L. Howes, D. R. Kaeli, P. Mistry, D. Schaa, "Chapter 13—OpenCL Profiling and Debugging", en *Heterogeneous Computing with Opencl*, Editado por B. R. Gaster, L. Howes, D R. KaeliPerhaad, M.D. Schaa, M. Kaufmann, Boston, 2013, pp. 243-261, ISBN 9780124058941.
- [16] NVIDIA, *NVIDIA CUDA Compute Unified Device Architecture*, 2014. Disponible en: [http://www.nvidia.com/object/cuda\\_home\\_new.html](http://www.nvidia.com/object/cuda_home_new.html)
- [17] E.W. Weisstein, *K-Means Clustering Algorithm*. 2014. Disponible en: <http://mathworld.wolfram.com/K-MeansClusteringAlgorithm.html>
- [18] M.I. Ángeles-Larrieta, A.M. Santillán-Gómez "Minería de datos: Concepto, características, estructura y aplicaciones", *Contaduría y Administración*, no. 190, p.p. 79-84, 2004.
- [19] Vallejos, S.J. "Minería de datos", *Facultad de Ciencias Exactas, Naturales y Agrimensura*, Argentina: Universidad Nacional del Nordeste, 2006.
- [20] P. Graham, *Parallel Computation Notes*, 1996. Disponible en: [http://www.cs.umanitoba.ca/~comp4510/notesDIR/COMP4510Models\\_2up.pdf](http://www.cs.umanitoba.ca/~comp4510/notesDIR/COMP4510Models_2up.pdf).



# Análisis de la Penetración de las Tecnologías de la Información y Comunicación (TIC) y su Influencia en la Reducción de la Brecha Digital en el Valle de Aburrá, Caso de Estudio sobre Internet

Analysis of the Penetration of Information-Communication Technologies (ICT) and its Influence on Bridging the Digital Gap into the Aburrá Valley's Population, An Internet Case of Study

**Gloria Amparo Rosales-Acevedo, MSc**

Facultad de Minas, Universidad Nacional de Colombia  
Medellín, Colombia  
garosalesa@unal.edu.co

**Sergio Botero-Botero, PhD.**

Facultad de Minas, Universidad Nacional de Colombia  
Medellín, Colombia  
sbotero@unal.edu.co

(Recibido el 30-10-2014. Aprobado el 15-12-2014)

**Resumen.** El presente artículo estudia la importancia del desarrollo de la penetración de Internet en un área geográfica. Para tal fin, se hace una descripción de cómo esta tecnología permite la disminución de la brecha digital, por medio del desarrollo económico. A partir del análisis teórico, se efectúa una revisión histórica del comportamiento de la penetración de Internet en la última década en el Valle de Aburrá (2003-2012), así como las variables de mayor influencia (de acuerdo con la literatura especializada). Con estas se desarrollan tres modelos de regresión lineal, en los cuales se identificaron las variables independientes relevantes. En el primero de los modelos, se toma como variable dependiente la penetración de conexiones de internet dedicadas; en el segundo y el tercero, se toma como variable dependiente el PIB per cápita del departamento de Antioquia y del Valle de Aburrá respectivamente.

**Palabras claves:** Brecha digital, desarrollo económico, penetración de internet dedicado.

**Abstract.** The paper shows the importance to develop Internet's penetration into a geographic area. This is achieved through a description of how such technologies aim to diminish the digital gap, by means of the economic development. Based upon theoretical analysis, there is a review of historic data on the Internet's penetration performance during the last decade (2003-2012) in the Aburrá Valley, including the most important variables (in accordance with specialized literature). Therefore, it is possible to perform three different linear regression models, making it possible to identify independent relevant variables. The first model takes into account a dependent variable related to dedicated Internet connections, whereas both second and third model define the dependent variable on the GDP per capita both in the department of Antioquia and in the Aburrá Valley.

**Keywords:** Digital Gap, Economic Development, Dedicated Internet Penetration.

**Citación de artículo, estilo IEEE:**

G.A. Rosales-Acevedo, S. Botero-Botero, "Análisis de la Penetración de las Tecnologías de la Información y Comunicación (TIC) y su Influencia en la Reducción de la Brecha Digital en el Valle de Aburrá, Caso Internet", *Lámpsakos*, N° 13, pp. 62-71, 2015.

**DOI:** <http://dx.doi.org/10.21501/21454086.1492>

## 1. INTRODUCCIÓN

El servicio de Internet como tecnología ha permitido desarrollar rápidamente la forma en que se ejecutan actividades sociales y económicas al acelerar el intercambio de información. Esta capacidad de descentralizar procesos facilita la optimización de actividades que generan oportunidades de eficiencia, productividad y efectividad.

A nivel mundial, existe preocupación por medir y documentar el desarrollo del Internet, dado que este es un acelerador de la economía que permite soportar diferentes actividades de tipo laboral, productivo y de entretenimiento. Por lo anterior, se ha generado un cambio en la forma de afrontar problemas de diversa índole: en el ámbito social, en la educación, la salud, el gobierno y el medio ambiente, por medio de aplicaciones electrónicas. En este orden de ideas, es importante ampliar su cobertura para la población en general (rural o urbana), en las velocidades adecuadas, para así garantizar la igualdad de accesibilidad a la información. Lo anterior constituye una motivación para asumir un trabajo de investigación que permita entender cuáles son los factores que inciden en el desarrollo de la penetración de Internet.

En esta investigación, se busca identificar las principales variables que influyen en la penetración de Internet dedicado en el Valle de Aburrá, que abarca las áreas urbanas y rurales del área metropolitana de la ciudad de Medellín y municipios aledaños (Colombia). Primero se presenta un marco conceptual de la relación entre la penetración de Internet y el crecimiento económico, luego se expone una revisión de la literatura sobre el tema y, finalmente, se plantea el desarrollo metodológico, donde se muestran tres modelos de regresión, los cuales identifican las principales variables relacionadas con la penetración de internet y el producto interno bruto. Al final del artículo, se efectúa un análisis de resultados, con sus respectivas conclusiones y recomendaciones.

## 2. ANTECEDENTES DE LA RELACIÓN ENTRE LA PENETRACIÓN DE INTERNET BANDA ANCHA Y EL CRECIMIENTO ECONÓMICO

La conectividad es el factor clave del flujo de información que define las economías modernas y es el elemento clave de la transformación en curso. Esta se entiende como el conjunto de cables de cobre, cables de fibra óptica y otros que permiten tener mu-

chos equipos conectados en red, lo cual optimiza los costos asociados, en vez de obtener la información a través de viajes físicos.

En resumen, se utiliza el término “conectividad” [1] para referirse a la totalidad de la interacción de la infraestructura de telecomunicaciones de una nación, la cual incluye: hardware, software, redes, internet banda ancha, computadores, redes avanzadas de datos corporativos y uso avanzado de servicios inalámbricos, sumando también las capacidades humanas, relacionadas con el uso de esta infraestructura.

Es ampliamente aceptado, de acuerdo a [1] que la mayoría de las economías modernas son de información y que una parte importante del crecimiento económico y de la productividad de un país desarrollado (por ejemplo: Europa, América del Norte y el Lejano Oriente Asiático) son impulsados por las tecnologías de información y las comunicaciones (TIC). Lo que es menos claro y reconocido es que en realidad todas las economías son de información; la libre circulación y disponibilidad de esta reduce las barreras a actividades económicas, estimulando el crecimiento y la productividad, incluso en las que normalmente no se consideran plenamente desarrolladas.

Al comparar los países de la OCDE con los países de América Latina y el Caribe, en 2009, los primeros alcanzaban una penetración en Internet banda ancha del 27% y los segundos del 6%. En Internet móvil llegaban al 47% y el 4% respectivamente [2]. Recuérdese que el término OCDE se refiere a la Organización para la Cooperación y el Desarrollo Económico y se constituye en la más importante organización de cooperación internacional, compuesta por 34 estados, cuyo objetivo es coordinar sus políticas económicas y sociales. Fue fundada en 1960 y su sede central se encuentra en París (Francia).

El aporte al desarrollo de la masificación del Internet banda ancha radica en la oportunidad de lograr una igualdad al acceso. Este solo se logra si se mejora la calidad de las redes, el marco normativo, políticas de uso y si se hace un claro enfoque en lineamientos hacia las regiones rurales[3]. El papel de la banda ancha en el cierre de la brecha digital es grande, al ser un condicionante de la competitividad de los países y de la inclusión social [2].

El término “brecha digital” [2] se refiere a la diferencia que existe entre sectores que tienen acceso a las herramientas tecnológicas y los que no lo tienen. Se ha

demostrado que con la revolución de la información se pueden reducir las diferencias entre los diferentes sectores de la sociedad, puesto que todos los habitantes pueden tener acceso a la misma información, mejorando su educación y formación.

Si las sociedades apoyan el desarrollo de internet, pero este tiene mayor penetración en las zonas urbanas que en las rurales, o en los sectores de mayores ingresos que en los de menores ingresos, se profundiza la exclusión social. Así que se deben hacer grandes esfuerzos para reducir la brecha digital y lograr la igualdad de oportunidades ofrecida en la revolución de la información.

En Colombia, de acuerdo a la Encuesta de Calidad de Vida del año 2013 [4], se puede observar la gran diferencia entre la penetración de Internet fijo de las cabeceras municipales y del resto de población; en 2012, la diferencia es de 33 puntos y en el 2013 es de 32.

De acuerdo a [5], se considera banda ancha en Colombia a las conexiones a Internet fijo con velocidad efectiva de bajada (Downstream) mayores o iguales a 1.024 Kbps o el Internet Móvil 3G y 4G.

Del total de suscriptores de Internet en Colombia, para el final del 2012, el 86% eran Internet banda ancha. De estos, el 62% (4.038.901 servicios) eran Internet fijo dedicado; esto quiere decir que se tiene una penetración del 8,7%, si se hace un cociente con la población total del país.

Si se lograra tener a la mayoría de colombianos conectados a Internet fijo dedicado de banda ancha (el cual permite la transmisión de grandes volúmenes de datos con posibilidades de uso de aplicaciones más avanzadas), se impactaría el crecimiento económico y desarrollo social al lograr la igualdad de oportunidades.

### **2.1. Revisión de estudios sobre el impacto de la banda ancha de internet sobre la economía**

Varios estudios han reportado el efecto que tiene la penetración de Internet banda ancha en el desarrollo económico. En la tabla No. 1, se presentan algunas referencias relacionadas con este efecto. Dicha tabla se construye con información recopilada por la UIT (Unión Internacional de Telecomunicaciones), orga-

nismo especializado de las Naciones Unidas para las tecnologías de la información y la comunicación – TIC.

Al revisar la literatura existente, se concluye en las investigaciones que el aumento en la penetración del servicio de Internet da un positivo incremento en la economía [10]. La relación entre penetración e impacto económico no es una relación lineal [11] porque la banda ancha podrá ser adoptada inicialmente por unas personas que la ven con gran beneficio y existirán unos adoptantes tardíos que la verán para lograr un menor beneficio. Cuando aumente la infraestructura y se alcance un nivel de masa crítica en la penetración, se verá el impacto en el crecimiento económico.

### **2.2. Relación entre penetración de internet y desarrollo económico**

Según se reportó en [8], en los países del OCDE aumentó el crecimiento económico a medida que mejoraba la penetración de banda ancha.

La banda ancha es un contribuyente al crecimiento económico en varios niveles [10]. En primer lugar, el despliegue de la banda ancha a través de empresas comerciales mejora la productividad al facilitar la adopción de procesos de negocio más eficientes. En segundo lugar, el amplio despliegue de la banda ancha acelera la innovación mediante la introducción de nuevas aplicaciones y servicios de consumo. En tercer lugar, la banda ancha lleva a una implementación funcional más eficiente de las empresas, por ejemplo, mediante la maximización de su alcance a las bolsas de trabajo.

El estudio del impacto de la banda ancha en el crecimiento económico abarca numerosos aspectos que van desde el impacto agregado sobre el crecimiento del PIB y el impacto diferencial de la banda ancha sobre el sector industrial hasta el aumento de las exportaciones, los cambios de demanda intermedia y la sustitución de importaciones.

De conformidad con la información presentada en [12], es necesario estimar la demanda de tecnología de banda ancha en América Latina y cuantificar el impacto macroeconómico de la tecnología de banda ancha en el empleo y la productividad. En la Tabla 2 se hace un resumen de los trabajos relevantes sobre los factores que impactan la penetración de banda ancha en esta zona geográfica.



**Tabla 1.** Resultados de la Investigación del impacto de Banda Ancha en el Crecimiento del PIB registrado por la UIT [10]

Alcance	Autores–Institución	Datos	Efecto
Estados Unidos	[6]– Brookings Institution	48 Estados de USA entre 2003 – 2005	Sin resultados estadísticamente significati- vos
	[7] – Ohio University	48 Estados de USA entre 2001 – 2005	Un aumento del 10% en penetración de banda ancha se asocia con 3,6% de aumento de la eficiencia del país.
OECD	[8] – Imperial College	22 países OECD entre 2002 – 2007	Un aumento de la banda ancha en la penetración de 10%. Se obtiene 0,25% en aumento del PIB.
66 Países con altos ingresos de la economía	[9]– Banco Mundial	66 países de grandes ingresos entre 1980 – 2002	Aumento del 10% en banda ancha produjo una penetración adicional 1,21 puntos porcentuales del crecimiento del PIB.
Países con bajos y medios ingresos de la economía	[9]– Banco Mundial	El resto de países 1980 – 2002	Aumento del 10% en penetración de banda ancha produjo un crecimiento adicional del PIB en un 1,38

**Tabla 2.** Variables Utilizadas en Estudios del Impacto de la Banda Ancha sobre el Desarrollo Económico en América Latina (Fuente: Elaboración de los Autores)

Autor	Nombre	Metodología	Variables Determinantes
Katz [12]	Estimating broadband demand and its economic impact in Latin America	Análisis de regresión lineal simple	Número de conexiones de banda ancha, penetración PIB de los países de la región en el 2008.
Cintel [13]	Medición de impacto del plan Vive Digital en Colombia y de la masificación de internet en la estrategia de Gobierno en Línea	El análisis realizado está basado en la construcción de modelos econométricos que permiten estimar el impacto económico de la banda ancha en Colombia en las tres áreas definidas y, posteriormente, proyectar el impacto esperado en el 2014	Conexiones Banda Ancha, Crecimiento Población, PIB
Callorda [14]	Desarrollo de banda ancha a nivel provincial en América Latina: determinantes del nivel de penetración	La metodología aplicada es el análisis econométrico de la información disponible para cada uno de los países mencionados, por separado y en conjunto.	Número de habitantes, el grado de urbanización, la magnitud del territorio y el Producto Bruto Interno (PBI) Per Cápita.
Galperin [15]	V. Las tarifas de banda ancha: benchmarking y análisis. Acelerando la revolución digital: banda ancha para América Latina y el Caribe	Cálculo IDBA, modelo simple de regresión con cinco factores identificados en la literatura.	<ul style="list-style-type: none"> <li>• PIB per cápita en dólares PPP (PBICAP)</li> <li>• Densidad de hogares por km2 (DENS-HH)</li> <li>• Porcentaje de población urbana (PO-PURB)</li> <li>• Porcentaje de la población entre 15 y 64 años (EDAD)</li> <li>• Un índice de educación (elaborado por Naciones Unidas) que captura la tasa de analfabetismo y la tasa de enrolamiento en los niveles primario, secundario y terciario (EDUC).</li> </ul>

El aumento de la difusión de Internet y de las tecnologías de la información y las comunicaciones mejora el proceso de conectividad, permitiendo el desarrollo económico, no solo en las grandes economías, sino en las que se encuentran en desarrollo, como América Latina.

Mientras que el número total de líneas de acceso de banda ancha en América Latina es de 26,8 millones y ha aumentado un 38% en los últimos años, la región aún tiene que crecer el número de líneas en un 41% (sumando 11 millones de líneas) para responder a las necesidades de la economía. Si eso llegara a ser alcanzado, se estima que el despliegue podría dar lugar al menos a 378.000 nuevos puestos de trabajo.

El análisis realizado en [13] está basado en la construcción de modelos econométricos que permiten estimar el impacto económico en Colombia y posteriormente proyectar el impacto esperado en el 2014. El resultado del estudio es que el crecimiento de las conexiones de banda ancha es la única variable independiente que logra explicar significativamente el crecimiento del PIB en todas las situaciones analizadas (tanto para los departamentos con alta como con baja penetración).

Por su parte, el estudio presentado en [14] permite relacionar los determinantes de las tasas de penetración de banda ancha a nivel provincial en América Latina. Para tal fin, se utilizan los datos trimestrales de Argentina (INDEC), Chile (SUBTEL) y de Colombia (MINTIC), entre el año 2008 y 2011.

En estos casos cabe aplicar el principio de subsidiaridad para que la intervención estatal fomente el crecimiento del número de conexiones de banda ancha. El autor [14] (para el caso de Colombia) estudia los departamentos con menor conectividad en Colombia (Guaviare, Guainía, Vaupés y Vichada), los cuales tienen un nivel de penetración menor al 1% entre sus habitantes. Asimismo, se efectúan observaciones en los departamentos colombianos con alta conectividad de internet (Bogotá -Distrito Capital-, Antioquia, Risaralda, Santander). Con dicha información ha sido posible ejecutar un modelo econométrico para diferentes departamentos de Colombia con los siguientes resultados:

1. Al aumentar en mil personas la cantidad de habitantes en el departamento, la cantidad de conexiones cada 100 personas mejora en un 0.0026; así también, al aumentar en un 1% la cantidad

de habitantes en zonas urbanas, la cantidad de conexiones cada 100 hogares aumenta en 0.2493. Ambos resultados hace evidente que el principal determinante del nivel de conectividad en Colombia son las causas sociodemográficas.

2. El nivel de PIB per cápita, como la superficie de los departamentos, parece no tener incidencia en el nivel de inclusión en el mundo digital.

Similares análisis se efectúan en [15], donde se aplican modelos econométricos en países latinoamericanos: Argentina, Belice, Bolivia, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Guyana, Honduras, Jamaica, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana, Surinam, Trinidad y Tobago, Uruguay y Venezuela, donde se validan las siguientes hipótesis:

- A mayor ingreso, mayor es la penetración de banda ancha –  $\alpha_1 > 0$
- A mayor densidad, mayor es la penetración de banda ancha –  $\alpha_2 > 0$
- A mayor porcentaje de población urbana, mayor es la penetración de banda ancha –  $\alpha_3 > 0$ .
- A mayor porcentaje de población entre 15 y 64 años, mayor es la penetración de banda ancha –  $\alpha_4 > 0$ .
- A mayor educación, mayor es la penetración de banda ancha –  $\alpha_5 > 0$

En general, los modelos econométricos revisados tienen en cuenta las siguientes variables:

- PIB per cápita en dólares PPP (PBICAP)
- Densidad de hogares por km<sup>2</sup> (DENS HH)
- Porcentaje de población urbana (POPURB)
- Porcentaje de la población entre 15 y 64 años (EDAD)
- Un índice de educación (elaborado por Naciones Unidas) que captura la tasa de analfabetismo y la tasa de enrolamiento en los niveles primario, secundario y terciario (EDUC).

Como resultado se puede comprobar que los signos de los coeficientes de los determinantes son los esperados y resultan significativos al 5%, a excepción de la variable edad.

Con los resultados anteriores, se procede a calcular el índice de desempeño en banda ancha (IDBA) [14], [15], el cual busca capturar esta dimensión que escapa a las comparaciones basadas en las tasas de penetración. El IDBA se presenta en la ecuación (2). Por ello, este indicador no se propone como alternativa, sino como complemento a los indicadores tradicionales de penetración.

La diferencia entre la penetración real y la predicha se define en este artículo, según las ecuaciones (1), (2)

$$PENET_i - PENET_{Pi} = \epsilon_i \quad (1)$$

$$PENET_i = \alpha_0 + \alpha_1 PBICAP_i + \alpha_2 DENS_{HH}_i + \alpha_3 POPUR_{Bi} + \alpha_4 EDAD_i + \alpha_5 EDUC_i + \epsilon_i$$

$$PENET_{Pi} = \alpha_0 + \alpha_1 PBICAP_i + \alpha_2 DENS_{HH}_i + \alpha_3 POPUR_{Bi} + \alpha_4 EDAD_i + \alpha_5 EDUC_i$$

$\epsilon_i$  es el estimador de  $\epsilon_i$ .

Para facilitar la interpretación de los resultados, se define al IDBA como:

$$IDBA = \epsilon_i / \max\{|\epsilon_i|\}. \quad (2)$$

Con este procedimiento, se obtienen valores del IDBA entre -1 y +1, donde los valores positivos indican un desempeño por encima de lo esperado, mientras que valores negativos apuntan a un desempeño por debajo de las expectativas. Un valor cercano a 0 indica que el país en cuestión tiene el desempeño esperable en función de su dotación de recursos económicos y sus características demográficas.

Para el Caso de Colombia el indicador IDBA es de -0,10 para el 2010, lo que significa que se encuentra por debajo de su potencial de desarrollo de sus características económicas y demográficas. Al hacer comparación con otros países de América Latina, se ven países como Nicaragua y Guatemala con baja penetración de Internet, pero con 0,05 como índice de banda ancha, lo que significa que tiene un adecuado nivel de penetración de acuerdo a sus condiciones económicas y demográficas. Por su parte,

Chile, país que lidera los rankings de penetración, también se encuentra por debajo de su nivel esperado, con un índice de -0,13.

### 3. DESARROLLO METODOLÓGICO: MODELOS DE REGRESIÓN Y SELECCIÓN DE VARIABLES

Este estudio está basado en las variables que afectan el incremento de la penetración de Internet fijo dedicado como primera opción de conectividad hacia la banda ancha. El espacio geográfico del presente estudio es el Valle de Aburrá, en el Departamento de Antioquia, Colombia. Dicha región incluye los municipios de Barbosa, Girardota, Copacabana, Bello, Medellín, Envigado, Itagüí, Sabaneta, La Estrella y Caldas.

Se limita la investigación a las observaciones de 10 años de penetración de Internet dedicado, ya que antes del 2003, las cifras que se llevaban por parte del Ministerio de las Tecnologías de la Información y las Comunicaciones en Colombia, MINTIC, son sobre utilización de Internet y no sobre conexiones fijas [16].

Se hizo una recopilación de datos con las posibles variables que, de acuerdo a la literatura revisada, podrían tener relación con los objetivos del presente estudio. Para identificar esta relación, se buscaron diferentes modelos de regresión; como resultado, se obtuvieron tres modelos, los cuales son descritos a continuación:

#### 3.1. Modelo de regresión 1

En una primera instancia de la investigación, se propuso un modelo de regresión lineal que contemple la Penetración de Internet Fijo Dedicado (Y) durante 10 años, desde el 2003 hasta el 2012, en el Valle de Aburrá, como variable que dependía de:

1. Producto Interno Bruto de Antioquia (PIB)
2. Densidad de hogares por KM<sup>2</sup> en el Valle de Aburrá (Dens)
3. Tasa de población urbana en el Valle de Aburrá (Pobl)
4. Proporción de la población del Valle de Aburrá entre 15 y 64 años (Edad)

**Tabla 3.** Coeficientes – Variable dependiente Penetración General Valle de Aburrá (Fuente: Los Autores)

Modelo	B	Sig.	Intervalo de confianza de 95,0% para B	
			Límite inferior	Límite superior
1 (Constante)	1,597	,097	-,416	3,610
Densidad_Hogares_ValleAburra	,001	,017	,000	,003
Porc Pob 15 a 64 General- VA	-,037	,057	-,075	,002
Matriculados Media	-,007	,023	-,013	-,001
Matriculados Superior	,006	,027	,001	,011

- 5. Tasa de Analfabetismo Antioquia (Analfabetismo)
- 6. Tasa bruta de personas matriculadas en educación primaria en Antioquia (EdPrim)
- 7. Tasa bruta de personas matriculadas en educación media en Antioquia (EdMed)
- 8. Tasa bruta de personas matriculadas en educación superior en Antioquia (EdSup)

Como resultado de la regresión, se obtuvo el siguiente modelo, con un coeficiente de correlación, R2, ajustado de 0.991:

$$(Y) \text{ Penetración de Internet dedicado fijo Valle de Aburrá} = 1.597 + 0.001 * \text{Dens} - 0.037 * \text{Edad} - 0.007 * \text{EdMed} + 0.006 * \text{EdSup} \quad (3)$$

Se pudo inferir que las variables estadísticamente significativas son: Dens, Edad, EdMed, EdSup, de acuerdo a la Tabla 3.

El modelo de regresión en (3) indica que:

- A. A medida que la densidad de hogares en el Valle de Aburrá aumenta en una unidad (un hogar por km2), se espera que la penetración de Internet dedicado fijo del Valle de Aburrá aumente en 0.001 unidades; esto siempre y cuando las demás variables permanezcan constantes.
- B. A medida que la proporción de edad entre 15 y 64 años del Valle de Aburrá aumenta en una unidad (1%), se espera que la penetración de Internet

dedicado fijo del Valle de Aburrá disminuya 0.037 unidades, mientras las demás variables permanecen constantes.

- C. A medida que la tasa bruta de matriculados en educación media de Antioquia aumenta en una unidad (1%), y si las demás variables en el modelo permanecen constantes, se espera que la penetración de Internet dedicado fijo en el Valle de Aburrá disminuya en 0.007 unidades.
- D. A medida que la tasa bruta de matriculados en educación superior de Antioquia aumenta en una unidad (1%), se espera que la penetración de Internet dedicado fijo del Valle de Aburrá aumente en 0.006 unidades, mientras las demás variables en el modelo permanecen constantes.

Cabe destacar que, inicialmente, la proporción de población entre 15-64 años presentó una significancia de 0.057 (no significativo al 5%, pero sí, por ejemplo, al 6%). Dicha variable se deja en el modelo debido a que su nivel de significancia está muy cerca al 5% tradicional (significancia de borde), por tener un valor-p dentro del rango [0.05-0.08], lo cual indica que de tener un tamaño de muestra mayor, podría resultar estadísticamente significativo.

### 3.2. Modelo de regresión 2

Según [10], en los casos de estudio presentados, también se puede tomar como variable dependiente el PIB per cápita.

**Tabla 4.** Coeficientes Variable Dependiente PIB Antioquia (Fuente: Elaboración Propia)

Modelo	B	Sig.	Intervalo de confianza de 95,0% para B	
			Límite inferior	Límite superior
1	(Constante)	,000	9038003,732	10496763,015
	Num usuarios General	,090	-,457	4,956
	Densidad+Tasa+Edad	,000	1205585,324	2454729,373

Por esta razón, en una segunda instancia de la investigación, se propuso observar si la penetración de internet fijo dedicado y las demás variables observadas influían significativamente en el desarrollo económico tanto del departamento de Antioquia como del Valle de Aburrá en sí, para lo cual se corrieron dos modelos de regresión lineal, teniendo como variable respuesta el PIB del Valle de Aburrá y el PIB de Antioquia.

Revisando las correlaciones entre las variables estudiadas, puede observarse una colinealidad casi perfecta (muy cercana a 1) entre algunas de las variables, por tanto, las variables densidad de hogares, tasa de población urbana y edad, se agrupan mediante componentes principales. El porcentaje de varianza explicada por la primera componente para la unión de estas variables es del 99.45%, es decir, se conserva gran parte de la información de las variables originales; a esta variable se le asignó el nombre de “Aspectos poblacionales”.

Usando el PIB de Antioquia (en pesos) y la cantidad de personas que acceden a Internet, junto con la variable unida de componentes principales (Aspectos poblacionales), se obtuvo el siguiente modelo ajustado, con un R2 ajustado igual a 0.992:

$$\text{PIB Antioquia} = 9'767.383,373 + 2,249 * \text{Num usuarios} + 1'830.157,349 * \text{Aspectos poblacionales} \quad (4)$$

Se obtienen los coeficientes reportados en la Tabla 4, indicando las siguientes conclusiones en el modelo (4):

1. En la medida en que la cantidad de usuarios con acceso a Internet en el Valle de Aburrá aumenta en una unidad, se espera que el PIB de Antioquia aumente en 2,249 pesos, mientras las demás variables permanecen constantes.
2. A medida que aumentan los aspectos poblacionales del Valle de Aburrá en un 1%, se espera que el PIB de Antioquia aumente en 1'830.157,349 pesos, mientras las demás variables permanecen constantes.

### 3.3. Modelo de regresión 3

Usando el PIB del Valle de Aburrá (en pesos) y la cantidad de personas que acceden a Internet, junto con la variable unida de componentes principales (aspectos poblacionales), se obtiene el siguiente modelo, con un R2 ajustado de 0.992:

$$\text{PIB Valle de Aburrá} = 6'573.449,011 + 1,514 * \text{Num usuarios} + 1'231.695,896 * \text{Aspectos poblacionales} \quad (5)$$

Se obtienen los coeficientes reportados en la Tabla 5, de donde se puede deducir el siguiente análisis del modelo (5):

1. A medida que la cantidad de usuarios que acceden a Internet en el Valle de Aburrá aumenta en una unidad, se espera que el PIB del Valle de Aburrá aumente en 1.514 pesos, mientras las demás variables permanecen constantes.

**Tabla 5.** Coeficientes Variable Dependiente PIB Valle de Aburrá (Fuente: Elaboración Propia)

Modelo	B	Sig.	Intervalo de confianza de 95,0% para B	
			Límite inferior	Límite superior
1 (Constante)	6573449,011	,000	6082576,513	7064321,510
1 Num usuarios General	1,514	,090	-,308	3,335
Densidad+Tasa+Edad	1231695,896	,000	811358,924	1652032,869

2. A medida que aumentan los aspectos poblacionales del Valle de Aburrá en un 1%, se espera que el PIB Valle de Aburrá aumente en 1'231.695,896 pesos, mientras las demás variables permanecen constantes.

Es importante aclarar que aunque la variable número de usuarios de Internet tiene una significancia de 0,09, no alcanza a estar dentro del rango de significancia de borde, por no tener un valor-p dentro del rango [0.05 – 0.08]; sin embargo, se deja dentro del modelo, pues por la cercanía con el límite, con un tamaño de muestra mayor, podría resultar estadísticamente significativo.

#### 4. CONCLUSIONES Y RECOMENDACIONES

*Esta investigación encuentra que para el Valle de Aburrá, el principal determinante del aumento del nivel de penetración de Internet fijo dedicado es la tasa bruta de matriculados en educación superior de Antioquia, seguida de la densidad de hogares por Km2; lo anterior refuerza el estudio [14] porque, de acuerdo a sus conclusiones, el principal determinante del nivel de penetración en Colombia son variables sociodemográficas como el nivel de habitantes, los cuales necesitan de acciones gubernamentales para mejorar la brecha digital.*

Se recomienda retomar el trabajo [17], en donde se identifican barreras claras para América Latina, como la accesibilidad; para esto, se deben revisar políticas frente a la competencia que bajen los precios de acceso. Otra herramienta para aumentar la demanda y reducir barreras económicas es la reducción de

impuestos asociados a los planes básicos de conectividad en hogares con bajos ingresos. También se podría impulsar el Internet móvil a través de la telefonía celular con bajos precios; para esta opción se requiere un papel muy proactivo del Gobierno en la estructura de oferta de banda ancha móvil.

Al revisar los modelos de regresión, en los cuales el PIB per cápita de Antioquia y el PIB per cápita del Valle de Aburrá son las variables dependientes, se puede concluir para ambos casos que el principal determinante para el crecimiento económico son las variables sociodemográficas definidas en este estudio, como aspectos poblacionales del Valle de Aburrá; la segunda variable es el nivel de usuarios de internet fijo dedicado en la misma zona.

Para trabajos futuros, como posibles profundizaciones de esta investigación, se propone:

Generar un estudio de la penetración de Internet móvil banda ancha como herramienta de interconexión de poblaciones sin conexión de fibra óptica.

Desarrollar un sistema de medición del Plan Vive Digital para las poblaciones rurales y municipales del departamento de Antioquia.

De acuerdo a esta investigación, para mejorar las limitaciones de significancia estadística en las variables que la literatura explica como determinantes en la penetración de Internet, se debe realizar una homologación en el tiempo de los datos históricos (desde el 2003) del Valle de Aburrá, de las variables de educación (Analfabetismo, tasa neta de escolaridad en la educación primaria, tasa neta de escolaridad en la educación media, tasa neta de escolaridad en

la educación superior), y de la cantidad de hogares y viviendas rurales y urbanas por municipio. Se recomienda utilizar metodología propuesta por [18].

Determinar adecuadamente los indicadores de educación de acuerdo a [17]. Esto ayudaría a afrontar la barrera de la habilidad, ya que se deben revisar las estructuras de los sistemas de educación existentes.

## REFERENCIAS

- [1] L. Waverman, N. Brooks, "Connectivity Scorecard 2009", 2009. Disponible: <http://www.connectivityscorecard.org/>.
- [2] CEPAL, "Las TIC para el crecimiento y la igualdad : Renovando las estrategias de la sociedad de la información," p. 113, 2010. Disponible: <http://www.cepal.org/cgi-bin/getProd.asp?xml=/socinfo/noticias/documentosdetrabajo/6/41716/P41716.xml&xsl=/socinfo/tpl/p38f.xsl&base=/elac2015/tpl/top-bottom.xsl> .
- [3] J. Benavides, F. Castro, "Impacto de las Tecnologías de la Información y las Comunicaciones (TIC) en el desarrollo y la competitividad del país", p. 225, 2011. Disponible: <http://www.repository.fedesarrollo.org.co/handle/11445/180>:
- [4] DANE, "Encuesta Nacional de Calidad de Vida 2014," 2015, Disponible: [http://www.dane.gov.co/files/investigaciones/condiciones\\_vida/calidad\\_vida/Presentacion\\_ECV\\_2014.pdf](http://www.dane.gov.co/files/investigaciones/condiciones_vida/calidad_vida/Presentacion_ECV_2014.pdf).
- [5] MINTIC, "Boletín trimestral de las TICs—Cifras cuarto semestre del 2012," 2013. Disponible: [http://colombiatic.mintic.gov.co/602/articulos-5550\\_archivo\\_pdf.pdf](http://colombiatic.mintic.gov.co/602/articulos-5550_archivo_pdf.pdf)
- [6] R. W. Crandall, "Broadband communications," Capítulo de Libro *The Handbook of Telecommunications Economics*. Editado por M. E. Cave, S. K. Majumdar, I. Vogelsang, vol. II, North-Holland, 2003. Disponible: <http://www.brookings.edu/~media/research/files/papers/2003/5/telecommunications-crandall/200305.pdf>
- [7] H. G. Thompson, "Broadband impacts on state GDP : Direct and indirect impacts," pp. 1–17. Disponible: [http://www.itu.int/wsis/stocktaking/docs/activities/1287145862/Ohio\\_University.pdf](http://www.itu.int/wsis/stocktaking/docs/activities/1287145862/Ohio_University.pdf)
- [8] P. Koutroumpis, "The economic impact of broadband on growth: A simultaneous approach," *Telecomm. Policy*, vol. 33, no. 9, pp. 471–485, 2009.
- [9] C. Qiang, "Economic impacts of broadband," *Increasing Impact*, pp. 35–50, 2009.
- [10] R. Katz, "The Impact of Broadband on the Economy: Research to date and Policy Issues. International telecommunication Union, The impact of Broadband on the Economy Broadband Series," *Telecommunication Dev. Sect.*, 2012.
- [11] M. A. Gillett, S., Lehr, W., and Osorio, C., & Sirbu, "Measuring Broadband's Economic Impact. Technical Report 99-07-13829, National Technical Assistance, Training, Research, and Evaluation Project.," pp. 1–53, 2006.
- [12] R. Katz, "Estimating broadband demand and its economic impact in Latin America," *Proc. 3rd ACORN-REDECOM ...*, pp. 1–20, 2009.
- [13] CINTEL, "Medición de impacto del Plan Vive Digital en Colombia y de la masificación de internet en la estrategia de Gobierno En Línea Adición N° 1 del Convenio Interadministrativo De Cooperación, no. 308 de 2011 Coordinación," pp. 1–59, 2011.
- [14] F. Callorda, "Desarrollo de banda ancha a nivel provincial en América Latina: Determinantes del nivel de penetración," *acorn-redecom.org*, pp. 17–18, 2012.
- [15] Ruzzier Galperin, "V. Las tarifas de banda ancha: *benchmarking* y análisis. *Acelerando la revolución digital: banda ancha para América Latina y el Caribe*," no. 2009, pp. 143-182, 2013.
- [16] CRT, "Las telecomunicaciones en Colombia," pp. 11–26, 2000. Disponible: [https://www.crcm.gov.co/uploads/images/files/cap01\\_las\\_telecomunicaciones\\_en\\_Colombia.pdf](https://www.crcm.gov.co/uploads/images/files/cap01_las_telecomunicaciones_en_Colombia.pdf)
- [17] Katz & Galperin, "*II. La brecha de demanda: determinantes y políticas públicas; Banda Ancha en América Latina: más allá de la conectividad*," *Nac. Unidas*, 2013.
- [18] R. Barro and J. Lee, "A new data set of educational attainment in the world, 1950-2010," *J. Dev. Econ.*, 2012. Disponible en: <http://www.nber.org/papers/w15902>.



# Incorporación de Proactividad a los Agentes en un Observatorio Tecnológico

## Proactivity Incorporation into Agents of a Technological Observatory

**Yahima Hadfeg-Fernández, MSc**

*Instituto Superior Politécnico José Antonio Echeverría  
La Habana, Cuba  
yhadfeg@ceis.cujae.edu.cu*

**Nelson Morales-Aguiar, Ing.**

*Instituto Superior Politécnico José Antonio Echeverría  
La Habana, Cuba  
nmorales@ceis.cujae.edu.cu*

**Mailyn Moreno-Espino, PhD.**

*Instituto Superior Politécnico José Antonio Echeverría  
La Habana, Cuba.  
my@ceis.cujae.edu.cu*

(Recibido el 30-10-2014. Aprobado el 15-12-2014)

**Resumen.** Uno de los paradigmas utilizados para desarrollar observatorios tecnológicos es el paradigma de agentes, debido a las múltiples funcionalidades que estos brindan. Una de las características más tentadoras que tienen los agentes es la capacidad de adquirir conocimiento, ya que pueden ser provistos de proactividad. De esta forma, pueden ejecutar mecanismos inteligentes para actuar en el entorno donde se encuentran y socializar con los demás agentes, incrementando sus prestaciones y acercándolos más a las necesidades reales de las personas. En este trabajo se presentan mecanismos para dotar de conocimiento a los agentes que forman parte de la arquitectura de un observatorio tecnológico. Esto se logró mediante el uso de métodos empíricos (encuesta), el motor de reglas Jadex–Rules y técnicas de minería de datos, apoyándose en la herramienta Knime para la extracción de reglas, permitiendo incorporarle un cierto grado de proactividad a los agentes del observatorio tecnológico.

**Palabras clave:** Agentes; minería de datos; motores de reglas; proactividad.

**Abstract** One of the paradigms used to develop Observation Technology corresponds to oriented agents, relating to multiple functions that they provide us. Furthermore, among the most enticing features, agents relate to their ability to acquire knowledge, since they are able to incorporate proactive features. Hence, they run intelligent mechanisms in order to act in the environment where they socialize with other agents, increasing their work usefulness and bringing them closer to the real needs of people. This paper presents mechanisms that provide knowledge to agents subscribed to a Technological Observatory using a Jadex rule engine. Both rules and data mining processes rely on KNIME tool, which allows extraction of new rules that can be later incorporated to proactivity agents in a Technological Observatory.

**Keywords:** Agents; Data Mining; Rule Engines; Proactivity

**Citación de artículo, estilo IEEE:**

Y. Hadfeg-Fernández, N. Morales-Aguiar, M. Moreno-Espino, “Incorporación de Proactividad a los Agentes en un Observatorio Tecnológico”, *Lámpsakos*, N° 13, pp. 72-79, 2015.

**DOI:** <http://dx.doi.org/10.21501/21454086.1369>



## 1. INTRODUCCIÓN

En la actualidad, los cambios causados en el entorno tecnológico demandan nuevas aproximaciones para el diseño, planificación e implementación de los procesos de innovación. En este sentido, la innovación debe ser concebida en términos de la generación, asimilación, transformación y difusión de conocimiento; por lo que es necesario que las organizaciones cuenten con novedosos sistemas de gestión de la información que les permitan transformar datos en conocimiento de valor estratégico para sus operaciones. En muchas ocasiones, estos novedosos sistemas de gestión se traducen en implementación sistemática de la vigilancia tecnológica [1], [2] y el diseño de nuevos servicios con minería y descubrimiento de conocimiento[3], [4].

La vigilancia tecnológica es un proceso organizado, selectivo y permanente que se encarga de captar información del exterior sobre ciencia y tecnología. Este proceso permite seleccionar la información, analizarla, difundirla y comunicarla, para lograr aprender de ella, con el objetivo de tomar decisiones con el menor riesgo posible y poder anticiparse a los cambios que pudieran ocurrir. Para llevar a cabo este proceso, es necesario que personas que sean expertos en el tema que se quiere vigilar lleguen a un consenso de lo que es más importante, para dar paso a las etapas de búsqueda de la información [5].

La vigilancia tecnológica ha propiciado la aparición de centros llamados Observatorios Tecnológicos[6], [7], los cuales permiten obtener información actualizada acerca de temas relacionados con la ciencia y la tecnología, nuevos descubrimientos, innovaciones y otras informaciones de carácter estratégico para una institución, sector o país. Estos espacios han sido pensados para la divulgación de información que favorezca el desarrollo tecnológico.

En el mundo nos podemos encontrar con una amplia variedad de sistemas y herramientas que nos proveen de información actualizada sobre un dominio específico, de manera integrada, este es el caso del proyecto SIGUE [8], Sistema Dinámico de Información como Observatorio Inteligente de Egresados de la Universidad Nacional Abierta y a Distancia (UNAD) [9]. Existen, además, sistemas dedicados al tratamiento de información centrada en dominios, como son los Observatorios Tecnológicos del Instituto de Tecnología Cerámica (ITC) [10], del Plástico [11] y de

la Soldadura [12]. Como sus respectivos nombres lo indican, cada uno es una referencia en el tema que procesan, dando a cada sector del mercado información tecnológica confiable e importante para los usuarios, de forma dinámica, periódica y actualizada.

Los observatorios tecnológicos son desarrollados con diferentes paradigmas [13], [14], uno de los utilizados es el paradigma de Agentes [15], debidos a las múltiples funcionalidades que estos brindan. Una de las características más tentadoras es la capacidad que tienen los agentes de adquirir conocimiento [16].

En [17] se muestra una propuesta de arquitectura basada en agentes, que implementa una metodología para la selección de una estrategia para la introducción de un producto.

La fuerza de desarrollo del Complejo de Investigaciones Tecnológicas Integradas está dirigida por el avance de varios programas. Estos están compuestos por subprogramas y estos, a su vez, por varios proyectos, enfocados a resolver cada uno un problema determinado. La clasificación de los proyectos está conformada por tres grupos que tienen en cuenta el producto que se desea obtener: Desarrollo de Sistema de Gestión, Desarrollo de Plataformas y Desarrollo de Componentes.

Actualmente, en el Complejo se desarrolla un Observatorio Tecnológico (OT), el cual está basado en la plataforma de desarrollo de agentes Jade [18], por las facilidades que brindan [19].

Los agentes inteligentes [19] pueden ejecutarse como sistemas tradicionales para funcionalidades determinadas; sin embargo, también pueden tener características proactivas [20], [21]. De esta forma, pueden ejecutar mecanismos inteligentes para actuar en el entorno donde se encuentran y socializar con los demás agentes, incrementando sus prestaciones y acercándolos más a las necesidades reales de las personas.

El proceso de dotar de inteligencia y comportamientos proactivos a un agente puede realizarse por disímiles vías, como los motores de reglas[22], [23].

En este trabajo se muestra cómo incorporar características proactivas a los agentes que sustentan el OT del Complejo de Investigación, mediante reglas de inferencias. Las reglas de inferencias serán obtenidas por varias vías, una de ellas es la entrevista

a clientes; y, la otra, la aplicación de técnicas de minería de datos a grandes volúmenes de información recolectada.

## 2. PROACTIVIDAD EN LOS AGENTES, MOTORES DE REGLAS

La proactividad es la actitud en que la persona asume el pleno control de su conducta vital de modo activo, lo que implica la toma de iniciativa en el desarrollo de acciones creativas y audaces para generar mejoras, haciendo prevalecer la libertad de elección sobre las circunstancias de la vida; implica asumir la responsabilidad de hacer que las cosas sucedan [24].

La proactividad en los agentes se puede implementar utilizando diferentes mecanismos, como son: las máquinas de estado finito, aprendizaje automático, reglas de inferencias [25], entre otras.

Las reglas, en general, son difíciles de programar y mantener en Java [26]. Un problema potencial puede ser la frecuencia de cambio de las reglas, donde aplicar un cambio en la concepción de una clase implica, por lo menos, modificar el código, compilar e instalar la aplicación, pudiendo llevar horas o días. Una forma de evitar este proceso es permitir encapsular la lógica de negocio o base de conocimientos y separarla del código aplicativo. Una manera efectiva de hacerlo es utilizando sistemas de gestión de reglas de negocio.

En su forma más simple, un motor de reglas de negocio está compuesto de tres elementos: un conjunto de reglas, la base de conocimientos (conocida como área de trabajo) y el procesador de reglas. Las reglas son sentencias de la forma IF-THEN, de tal manera que si se cumplen todas las condiciones del IF se ejecutan todas las acciones del THEN. El motor utilizará la base de conocimientos para decidir que reglas deben activarse.

### 2.1. Obtención de reglas para el Observatorio Tecnológico

Para obtener las reglas que se implementarán, se utilizaron diferentes métodos y procedimientos: primero, se entrevistaron profesores de una muestra seleccionada; además, se utilizó la minería de datos [27] para obtener más reglas de interés.

Para realizar las entrevistas, se escogió una muestra de 38 personas que desempeñan el rol de profesores/investigador o especialistas del Complejo, con años de experiencia en la búsqueda de documentación científica. A esta muestra se le realizaron las siguientes preguntas:

- ¿Qué aspectos relacionados con el proceso de búsqueda de información considera usted que pudiera realizarse sin su supervisión?
- ¿Qué sugiere usted agregar al sistema para mejorar las búsquedas realizadas por los usuarios?

Se realizó la siguiente pregunta a 31 personas con mayor dominio del tema:

- ¿Qué funcionalidades considera usted que pudieran agregárseles al sistema en forma de reglas?

A continuación se muestran los resultados que arrojó la entrevista.

Para la pregunta 1, cada una de las personas entrevistadas aportó un conjunto de aspectos, para un total de 13.

La pregunta 2 arrojó 21 sugerencias.

En la tabla 1 se detallan para cada una de las preguntas, los aspectos y sugerencias obtenidos en el proceso.

**Tabla 1.** Resultados de las preguntas de las entrevistas

Preguntas de la entrevista	Aspectos a tener en cuenta
¿Qué aspectos relacionados con el proceso de búsqueda de información considera usted que pudiera realizarse sin su supervisión?	<p>Realizar búsquedas cada determinado tiempo.</p> <p>Recibir notificación de documentos encontrados que no se tenga en el repositorio personal del usuario.</p> <p>Mantener al usuario informado de algún cambio en los documentos de interés.</p> <p>Tener una base actualizada de documentos.</p> <p>Definir temáticas afines con sus intereses.</p> <p>Identificar usuarios con información común.</p> <p>Compartir con otros usuarios documentos que sean de interés común.</p> <p>Conocer el nivel de conocimiento de diferentes usuarios.</p> <p>Relacionar documentos, ya sean pertenecientes a la misma temática o a temáticas diferentes.</p> <p>Declarar documentos importantes de un tema.</p> <p>Actualizar documentos con versiones más recientes.</p> <p>Obtener una base de recursos.</p> <p>Dar relevancia a documentos.</p>
¿Qué sugiere usted agregar al sistema, para mejorar las búsquedas realizadas por los usuarios?	<p>Devolución de información deseada.</p> <p>Desglose de los materiales por tipo.</p> <p>Tener en cuenta las preferencias del usuario.</p> <p>Que se haga uso del perfil de usuario.</p> <p>Reducción de las palabras claves al campo de acción.</p> <p>Tener un perfil de usuario con la calidad requerida.</p> <p>Que no se repita el mismo documento que fue buscado con anterioridad.</p> <p>Definir un organigrama.</p> <p>Retroalimentación del sistema en dependencia de las búsquedas.</p> <p>Incluir palabras para descartar las búsquedas en determinado contexto.</p> <p>Modificar las palabras claves en dependencia del ambiente.</p> <p>Realizar un análisis semántico de las palabras claves.</p> <p>Crear niveles entre usuarios.</p> <p>Refinar las búsquedas, descartando o disminuyendo la cantidad de documentos a encontrar.</p> <p>Filtrar la búsqueda para reducir el número de documentos enviados.</p> <p>Sugerir el contexto en que se busquen las palabras claves.</p> <p>Sugerir palabras a descartar.</p> <p>Declarar calidad de los documentos enviados al usuario.</p> <p>Mantener al usuario informado de los nuevos documentos que contengan palabras claves cercanas a su interés.</p> <p>Permitir que el usuario sugiera un tiempo de búsqueda de sus palabras claves y la frecuencia con que quiera que aparezcan.</p> <p>Definir ontologías de dominios que los documentos que se devuelvan estén en el contexto deseado.</p>

La última pregunta de la entrevista solo fue aplicada a una muestra de 27 profesores, por ser estos los que más dominio tienen en el tema. De esta pregunta, se obtuvo un total de 5 reglas que se muestran a continuación:

1. Si dos o más usuarios coinciden en la búsqueda de palabras claves, ponerlos en contacto a través de correo electrónico para que sepan que buscan información sobre palabras claves en común.
2. Declarar como relevante un recurso que se haya obtenido muchas veces por los usuarios en sus búsquedas realizadas.
3. Si un recurso es el más devuelto a usuarios que pertenecen a un mismo proyecto, declararlo como relevante para el proyecto.
4. Cuando entre una persona nueva al proyecto, enviarle los recursos declarados como relevantes en dicho proyecto.
5. Si un usuario se declara seguidor de otro, enviarle los recursos recibidos por el usuario que él sigue.

La Fig. 1 muestra los conceptos y entidades que describen el dominio donde se impactará con esta investigación, así como sus relaciones.

Del proceso de realizar la entrevista se obtuvieron reglas a implementar, las cuales serán complementadas con las que se obtengan de aplicar minería de datos a las trazas del OT.

## 2.2. Minería de datos

Para incorporar características proactivas en los agentes del OT, se hace uso de la herramienta KNIME [28] y de técnicas de asociación, buscando reglas que describan las relaciones entre los atributos de las tablas de las bases de datos del OT y utilizando la metodología CRISP-DM [29].

La tabla “trace” de la base de datos del OT fue seleccionada como candidata debido a que es esta tabla la que presenta mayor volumen de datos en comparación con las demás, pues es aquí donde queda registrado todo lo que ocurre en el sistema.

Los datos colectados reflejan las búsquedas de los agentes personales que representan a cada usuario del Observatorio Tecnológico. Se tienen registradas 18584 instancias en la tabla “trace”, que es donde se registra todo el movimiento en el sistema.

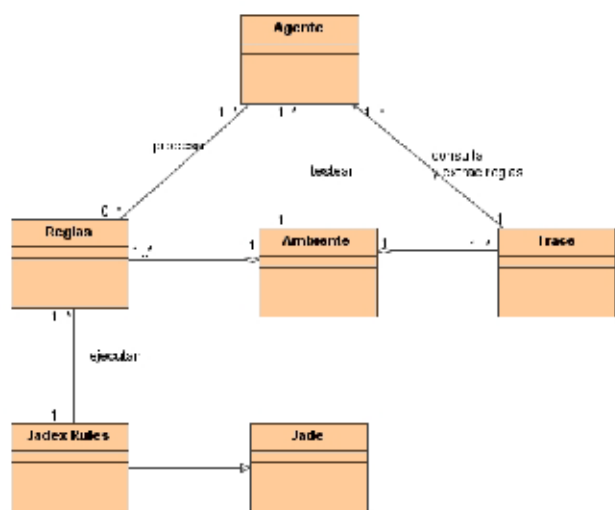


Fig. 1. Modelo de Dominio.

Esta tabla (*trace*) está compuesta por un conjunto de 23 atributos, de ellos 13 son de tipo *String*, 1 *long*, 5 *int*, 1 *Calendar*, 1 *boolean* y 2 *float*. Con el objetivo de obtener varias vistas minables, se decidió dividir la información de la tabla a través del atributo “*performative*”, el cual es de tipo *String*, describe el tipo de mensaje y puede tomar los valores de: INFORM o REQUEST.

Los resultados más importantes de la correspondencia con el valor INFORM de la “*performative*” son los siguientes: El agente que más envía información es el agente “FTP\_CITI”, representando un 63% de todos los agentes que envían información.

Para obtener las direcciones de los recursos que más devuelven información, se hace necesario realizar transformaciones, así el campo “*resource\_path*”, que es de tipo *string* y describe la dirección completa donde se guarda el recurso, queda dividido y agrupado en dependencia del lugar de origen.

A partir de esta transformación, se obtienen seis direcciones, donde la dirección del FTP del CITI es de donde se encuentran los recursos que buscan los usuarios, representando el 63 % de las direcciones.

Para el caso de la segunda vista minable, se obtuvieron los siguientes resultados:

El 20% de las palabras que se buscan corresponde con la palabra clave “web service”.

El 21% de las solicitudes de las búsquedas, a través de las palabras claves, la realiza el usuario que corresponde al agente personal “AP\_1060”.

Para definir reglas de asociación o de dependencia entre los datos y obtener un conjunto de estas, se realizaron 3 experimentos sobre la tabla “*trace*”.

Los experimentos 1 y 2 fueron realizados con la vista minable 1, mientras que para el experimento 3 se utilizó la segunda vista minable; estos se describen en la tabla 2.

**Experimento 1:** Se obtuvo un total de 23 reglas de asociación, con una confianza del 100%, que el agente “agent\_sender: FTP\_CITI” busca en el repositorio que tiene como dirección local FTP CITI.

Con una confianza del 100%, se tiene que la dirección FTP CITI devuelve a los agentes FTP\_CITI los recursos del sistema.

**Experimento 2:** Los resultados de este experimento fueron muy parecidos al experimento anterior. El aumento de la confianza a un 80% y el soporte a un 40% trajo como consecuencia que disminuyera el número de reglas, pero las obtenidas son iguales a las del experimento anterior.

**Experimento 3:** Debido a la calidad que tienen los datos este experimento no arrojó reglas para el sistema.

### 3. RESULTADOS Y DISCUSIÓN

A partir del análisis realizado, se identificaron cuatro reglas, implementables en el motor de regla Jadex-Rules.

**Regla 1:** Enviar correos a usuarios que buscan información con palabras claves iguales.

Para poder ejecutar esta regla, los usuarios deben haber introducido en el perfil de usuario palabras claves. La conjunción de estas palabras entre diferentes usuarios no debe ser nula. A continuación un ejemplo donde varios usuarios de diferentes programas y proyectos llenan la información de su perfil principal; todos los usuarios que son objeto de análisis tienen en común a “web service” como palabra clave.

**Tabla 2.** Parámetros que se describen en los experimentos 1,2 y 3.

Parámetros	Experimento 1 y 3	Experimento 2
Minimum set size	2	3
Minimumsupport	0.3	0.4
Minimum rule confidence	0.7	0.8

Cuando se ejecuta esta regla, el motor busca en la base de datos todos aquellos usuarios que solicitan información con palabras claves iguales y estos reciben un correo donde se les informa que hay en el sistema otros usuarios que buscan la misma palabra clave.

**Regla 2:** Declarar el (los) recursos(s) como relevante(s).

La ejecución de esta regla permite que, a partir de los resultados de las búsquedas realizadas, se declaren recursos relevantes y, de esta forma, tener una valoración de los mismos.

Los usuarios del sistema pueden evaluar un recurso mediante el perfil del usuario (relevancia manual), dando como máximo un total de 5 puntos en dependencia de la utilidad que estos le vean para el trabajo que desempeñan.

Esta regla incorpora un valor numérico (relevancia automática) que, sumado con la valoración del usuario, le permite al recurso tener una valoración completa. Se dice que un recurso está evaluado completamente cuando se le ha otorgado la relevancia manual (otorgada por el usuario) y la relevancia automática (calculada por el sistema). Para poder ejecutar esta regla es necesario que existan en la base de datos recursos devueltos por el sistema.

El valor de la relevancia automática de estos recursos se obtiene a partir de aplicar la siguiente fórmula:

Como máximo, cada recurso tendrá una relevancia con un valor de 10 puntos.

$$\text{Relevancia Automática} = \frac{\text{Cantidad de documentos por usuarios}}{\text{Total de usuarios del proyecto}} * 5$$

Si se supone que se tienen dos proyectos, como se muestra en la Fig. 2, que a su vez tienen tres usuarios y para los primeros dos usuarios de cada proyecto tienen asignados dos recursos iguales y uno diferente:

- |   |  |
|---|--|
| <p>1. Observatorio Tecnológico</p> <ul style="list-style-type: none"> <li>• Alternan             <ul style="list-style-type: none"> <li>○ Documento 1</li> <li>○ Documento 2</li> <li>○ Documento 3</li> </ul> </li> <li>• Mailyn             <ul style="list-style-type: none"> <li>○ Documento 1</li> <li>○ Documento 2</li> <li>○ Documento 4</li> </ul> </li> <li>• Leandro             <ul style="list-style-type: none"> <li>○ Documento 5.</li> <li>○ Documento 6</li> </ul> </li> </ul> | <p>2. Multimedia.</p> <ul style="list-style-type: none"> <li>• Nelson             <ul style="list-style-type: none"> <li>○ Documento 7</li> <li>○ Documento 8</li> <li>○ Documento 9</li> </ul> </li> <li>• Yahima             <ul style="list-style-type: none"> <li>○ Documento 7</li> <li>○ Documento 8</li> <li>○ Documento 10</li> </ul> </li> <li>• Nayma             <ul style="list-style-type: none"> <li>○ Documento 11</li> <li>○ Documento 12</li> </ul> </li> </ul> |
|---|--|

**Fig.2.** Ejemplo de programas con usuarios.

Aplicando la fórmula para el primer proyecto, quedaría de la siguiente forma:

$$\text{Relevancia Automática Documento 1} = (2/3) * 5 = 3,33.$$

Si el documento fue evaluado por el usuario con 5 puntos, sumado al resultado anterior, este tendrá una relevancia final de 8,33 puntos y este valor queda registrado.

En caso de que este documento no contenga evaluación por parte del usuario, quedará registrado el valor de la evaluación dado por el sistema, que para ese ejemplo sería 3,3 puntos.

**Regla 3:** Enviar el o los recurso(s) más relevantes de un proyecto a un usuario del sistema.

A partir de ejecutar la regla anterior, los recursos que devuelve el sistema tienen una valoración y, en acuerdo con los miembros del grupo de investigación, se decide a partir de qué puntuación un recurso es relevante para un proyecto.

Un recurso se considera relevante cuando su evaluación es mayor o igual que 5.0 puntos, debido a que este valor corresponde con la mitad de ambas partes (valoración manual y valoración automática).

**Regla 4:** Buscar el o los recurso(s) que se hayan obtenido de realizar búsquedas con diferentes palabras claves.

Cuando se ejecuta esta regla, se obtienen los recursos que fueron devueltos después de haber realizado búsquedas con diferentes palabras claves.

Por ejemplo, un usuario desea buscar información con la palabra clave “web services” y “bpm”, y otro usuario realiza búsquedas con “soap” y “agentes”. Del resultado de estas búsquedas se obtuvo el documento X, que fue devuelto porque contenía las palabras “web services” y “soap”.

A partir de esta información, se le envía a los usuarios que se les devolvió el documento X un correo informándoles que dicho documento fue resultado de una búsqueda con palabras claves diferentes a las que el empleó.

Esta regla es de interés para los usuarios del sistema, ya que mediante ella se pueden identificar temas adyacentes.

#### 4. TRABAJOS FUTUROS

En aras de mejorar los resultados de esta investigación, se recomienda:

Aumentar los datos registrados en la base de datos para obtener información con mayor grado de relevancia.

Incorporar mecanismos para engranar automáticamente las reglas obtenidas del proceso de minería de datos con el motor de regla Jadex-Rules.

Permitir mediante la herramienta que sean los jefes de proyectos los que determinen el grado de relevancia que debe tener un documento para que sea clasificado como relevante para el proyecto.

#### 5. CONCLUSIONES

Al concluir este trabajo, se definió un conjunto de intereses de los usuarios del Observatorio Tecnológico, que fueron planteados en formas de reglas. Las reglas son resultado de un descubrimiento de conocimiento apoyado por el proceso de minería de datos y las encuestas realizadas al personal del Centro. Los agentes de software del Observatorio Tecnológico adquirieron inteligencia a partir del motor de reglas Jadex-Rules y la base de conocimiento, quedando dotados de un comportamiento proactivo.

#### REFERENCIAS

- [1] M. M. Delgado-Fernández, M. Infante-Abreu, Y. Abreu-Lee, O. Infante-Pérez, A. Díaz-Batista, J. Martínez-Moreno, “Vigilancia tecnológica en una universidad de ciencias técnica”, *Ingeniería Industrial*, vol. 32, no. 1, pp. 69-75, 2011.
- [2] Y. Abreu-Lee, M.B. Infante-Abreu, T. Delgado-Fernández, M. Delgado-Fernández, “Modelo de vigilancia tecnológica apoyado por recomendaciones basadas en el filtrado colaborativo”, *Ingeniería Industrial*, vol 34, no. 2, pp. 167-177, 2013.
- [3] A. Cobo, R. Rocha, M. Alonso, “Descubrimiento de conocimiento en repositorios documentales mediante técnicas de minería de texto y swarm intelligence”, *Revista Electrónica de Comunicaciones y Trabajo de Asepuma*, vol., no. 10, pp. 105-124, 2009.
- [4] E. Durán, R. Costaguda, “Minería de datos para descubrir estilos de aprendizaje”. *Revista Iberoamericana de Educación*, vol. 42, no. 2, pp. 1-10, 2007.
- [5] Fundación para el Conocimiento. “Vigilancia Tecnológica”, Madrid (España), 2009, Disponible: <http://www.madrimasd.org/vigTecnologica/>.
- [6] M. Moreno-Espino, A. Carrasco-Bustamante, A. Rosete-Suárez, M. D. Delgado-Dapena. “Apoyo a la toma de decisiones en un Observatorio Tecnológico incorporando proactividad”. *Ingeniería Industrial*, vol. 34, no. 3, pp. 296-306, 2013.
- [7] A. Díez-Manjarrés, C. Angulo, “Análisis para la creación de un observatorio tecnológico”. Universidad Politécnica de Cataluña, Barcelona (España), 2011, Disponible: <http://bit.ly/1K06bty>
- [8] UNAD, “Investigación Universidad Nacional Abierta y a Distancia”. 2013, Disponible: [http://www.unad.edu.co/index.php?option=com\\_content&view=category&layout=blog&id=94&Itemid=719](http://www.unad.edu.co/index.php?option=com_content&view=category&layout=blog&id=94&Itemid=719)
- [9] UNAD, “Observatorio Colombiano. Universidad Nacional Abierta y a Distancia”, 2010, Disponible: <http://www.unad.edu.co/>.
- [10] ITC, “Observatorio Tecnológico”, 2014, Disponible: <http://www.observatoriotecnologicoceramico.es/itc/opencms/itc/es/>.

- [11] AIMPLAS. Observatorio del Plástico, 2010, Disponible: <http://www.observatorioplastico.com>.
- [12] AIMEN Centro Tecnológico, "Observatorio Tecnológico de la Soldadura", 2010, Disponible: [http://www.obtesol.es/index.php?option=com\\_content&task=view&id=5&Itemid=35](http://www.obtesol.es/index.php?option=com_content&task=view&id=5&Itemid=35).
- [13] J.J. Gómez-Sanz, "Metodologías para el desarrollo de sistemas multi-agente". *Revista Iberoamericana de Inteligencia Artificial*, vol. 7, no. 18, pp. 51-63, 2003,.
- [14] S. R. Schach&L. P. Rosales, "Análisis y diseño orientado a objetos con UML y el proceso unificado". Vol. 1. México: McGraw-Hill, 458 p. 2005.
- [15] Y. Lores-de-Feria, J. Fuentes, N. Elieser; J.C. Bayeux-Iglesias, A. Rosete-Suárez, "Requisitos y patrones: Retos para la ingeniería de software orientada a agentes", *Ingeniería Industrial*, vol. 27, no. 2-3, pp. 29-32, 2006.
- [16] E. M. Sánchez-Vila, M. Lama-Penín, "Técnicas de la inteligencia artificial aplicadas a la educación". *Revista Iberoamericana de Inteligencia Artificial*, vol. 11, no. 33, pp. 7-12, 2007.
- [17] P. M. N.Matsatsinis, V.Psomatakis, N.Spanoudakis, "An Agent-Based System for Products Penetration Strategy Selection", vol. 17, Taylor & Francis Inc., p. 24, 2003.
- [18] F. Bellifemine, "Jade programmere's guide", vol. 3. Boston, p. 61, 2002.
- [19] M. Wooldridge, "An Introduction to MultiAgent Systems". Great Britain: John Wiley & Sons, p. 453, 2009.
- [20] A. López-Salazar, "La Proactividad empresarial como elemento de competitividad". *Ra Ximhai, Revista de Sociedad, Cultura y Desarrollo Sustentable*, vol 6, no. 2, pp. 303-312, 2010.
- [21] J. Yáñez, P. Gaete, T. Harcha, W. Kühne, V. Leiva, P. Vergara., "Hacia una metateoría constructivista cognitiva de la psicoterapia". *Revista de Psicología*, vol. 10, no. 1, pp. 97- 110, 2001.
- [22] Jadex-Rule-Engine, "User Guide.01 Introduction: WebHome". 2010, Disponible: <http://jadex-rules.informatik.uni-hamburg.de/xwiki/bin/view/User+Guide/01+Introduction>.
- [23] M. Bali, "Drools JBoss Rules 5. X Developer's Guide", *Packt Publishing Ltd, Birmingham (R.U)* p. 338, 2013.
- [24] V. E. Frankl, "Man's Search for Meaning", *Herder S.A, Barcelona (España)*, p. 137, 1991.
- [25] S. Casas, C. Marcos. "Exploración de Reglas de Inferencia para Automatizar la Refactorización Aspectual". *II Latin American Workshop on Aspect-Oriented Software Development*, Campinas, Brasil, 2008.
- [26] Wordpress. "¿Cuándo utilizar un motor de reglas?" *Crónicas de un Arquitecto de Software*, 2009, Disponible: <http://dsolis.wordpress.com/2009/06/18/%C2%BFcuando-utilizar-un-motor-de-reglas/>.
- [27] J. Aguilar, J. Altamiranda, "Conceptos sobre minería web". *Revista Gerencia Tecnología Informática*, vol. 3, no. 7, 2006
- [28] M. R. Berthold. "KNIME: The Konstanz Information Miner". Vol. 1. New York: Springer Berlin Heidelberg. 2008.
- [29] A. Azevedo, M.F. Santos, "KDD, SEMMA and CRISP-DM: a parallel overview" *IADIS European Conference Data Mining*. Porto, Portugal, pp. 182-185, 2008. ISBN 978-972-8924-63-8

# REFLEXIÓN ANÁLITICA E INTERPRETATIVA





# Modelo de Simulación de Estrategias de Inversión para Papicultores Colombianos

## Simulation Model of Investment Strategies for Colombian Potato Farmers

**Víctor Daniel Gil-Vera, MSc.**

*Facultad de Ingenierías y Arquitectura  
Fundación Universitaria Luis Amigó  
Medellín, Colombia  
victor.gilve@amigo.edu.co*

(Recibido el 20-11-2014. Aprobado el 10-12-2014)

**Resumen.** En Colombia, la papa es un producto básico cuyo precio depende principalmente de la oferta y la demanda locales. El sector papicultor colombiano presenta grandes incertidumbres sobre los beneficios netos que se pueden generar en el corto, mediano y largo plazo, debido a la alta volatilidad de los precios del tubérculo. Este trabajo presenta un modelo de simulación de estrategias de inversión para papicultores colombianos desarrollado en la plataforma de simulación Vensim. El modelo consideró diferentes variables, entre las que se destacan el precio, el costo de la mano de obra, inversiones en nuevas tecnologías, entre otras. Con el desarrollo de este trabajo, se concluye que las utilidades de los productores son altamente sensibles a tres factores: precio, costos de producción e inflación.

**Palabras clave:** Dinámica de sistemas, estrategias de inversión, modelos de simulación, ciclo de balance.

**Abstract.** In Colombia, the potato is considered as a commodity whose price depends mainly on the local supply and demand. The Colombian potato's segment has large uncertainties on their benefits for either short, medium or long term. This is due to the high volatility of prices of the commodity. This paper shows a simulation model to guide investment strategies for different kinds of potato crops in Colombia, keeping in mind different variables such as price, cost and investment in new technologies. Hence, conclusions on potato crops' profits are then formulated as they are highly sensitive to three factors: price, production costs and inflation.

**Keywords:** System dynamics, investment strategies, simulation models, balance cycle.

**Citación de artículo, estilo IEEE:**

V.D. Gil-Vera, "Modelo de Simulación de Estrategias de Inversión para Papicultores Colombianos", *Lámpsakos*, N° 13, pp. 81-87, 2015.

DOI: <http://dx.doi.org/10.21501/21454086.1500>

## 1. INTRODUCCIÓN

En Colombia, la producción de papa representa la principal actividad agrícola desarrollada en zonas frías, y es el sustento de más de 100.000 familias [1]. La producción de la papa es una actividad caracterizada por una incipiente infraestructura tecnológica y un escaso nivel de gestión agroempresarial, cuya producción exige elevados costos y su rentabilidad depende de precios de mercado altamente fluctuantes [2].

En este artículo, se presenta un modelo de simulación de estrategias de inversión para papicultores colombianos, el cual considera diferentes variables que afectan, directa o indirectamente, sus utilidades, tales como el precio, el costo de la mano de obra y fertilización, la inversión en nuevas tecnologías, entre otras. En el sector papicultor en Colombia, se presentan grandes incertidumbres sobre los beneficios netos que se pueden generar en el corto, mediano y largo plazo.

No existe una fórmula que permita calcular con exactitud las ganancias netas de un papicultor, pero existen herramientas que permiten simular, con una exactitud aceptable, dichos valores. Debido a lo anterior, la aplicación de la dinámica de sistemas en este sector es una de las mejores alternativas para predecir y conocer las utilidades netas que pueden obtener los productores colombianos en el corto, mediano y largo plazo. Vensim fue la plataforma de simulación utilizada para la creación del modelo.

Este trabajo se divide de la siguiente forma. En primer lugar, se presenta una caracterización general del sector, la descripción del problema y del modelo desarrollado (diagrama causal, flujos y niveles), los resultados obtenidos y las políticas que ayudarían a mejorar las condiciones del problema.

## 2. CARACTERIZACIÓN DEL SECTOR

La papa es un producto de alto consumo, tiene un peso importante sobre la canasta familiar y un papel importante en la definición del índice general de precios de la economía [3]. En los últimos años, la papa en Colombia ha tenido producción en aumento, debido a que existe gran cantidad de cultivos. Las variedades más utilizadas son: tuquerreña o sabanera, parda pastusa, pastusa suprema, rubí, salentina, carriza, diacol capiro-r12, ICA única, ICA nevada, ICA

Nariño, milenio-1, diacol Monserrate y la papa criolla [3]. En la Figura 1, se presentan los datos de producción de papa en Colombia en el período 2008–2014.

En el país se cultiva papa en once departamentos, pero solo en cuatro de ellos (Cundinamarca, Boyacá, Nariño y Antioquia) se concentra alrededor del 90% del área y la producción. Cabe destacar que Cundinamarca y Boyacá son los dos departamentos donde se encuentran concentradas las áreas con mayor producción de papa del país. En la Figura 1, se observa que el comportamiento de estos dos departamentos fue constante durante el periodo 2008–2014. Los valores de producción se mantuvieron entre 2.490.800–2.788.050 Ton/Año, encontrándose una mayor producción durante el año 2012, equivalente a 2.788.050 Ton [3].

En Colombia, los pequeños productores, es decir, aquellos que siembran menos de tres hectáreas, constituyen el 95% de los cultivadores, quienes producen alrededor del 45% del total de la producción. En todo el territorio, se siembran en promedio 130 mil hectáreas y se producen cerca de 2,8 millones de toneladas de papa al año [4], dependiendo del régimen climático y la variación en los precios pagados al productor [5].

## 3. DESCRIPCIÓN DEL PROBLEMA

La alta volatilidad de los precios de la papa afecta la calidad de vida de los papicultores colombianos y dificulta la toma de decisiones de inversión. Actualmente, estos no cuentan con herramientas que les permitan analizar esquemas estratégicos, así como también analizar, en el mediano y largo plazo, el impacto de estas estrategias sobre sus ingresos, como lo son la implementación de fertilizantes, renovación de tierras, inversiones en tecnología, entre otras.

### 3.1. Diagrama causal

Cuando los papicultores buscan incrementar la producción a través del uso de fertilizantes, con la renovación de tierras e inversiones en tecnología, deben destinar una cantidad de capital considerable que puede afectar su liquidez. Al implementar las estrategias mencionadas anteriormente, se incrementa la producción de papa (ciclo de refuerzo); sin embargo, esto viene acompañado de un incremento en los costos generales (ciclo de balance) [6].

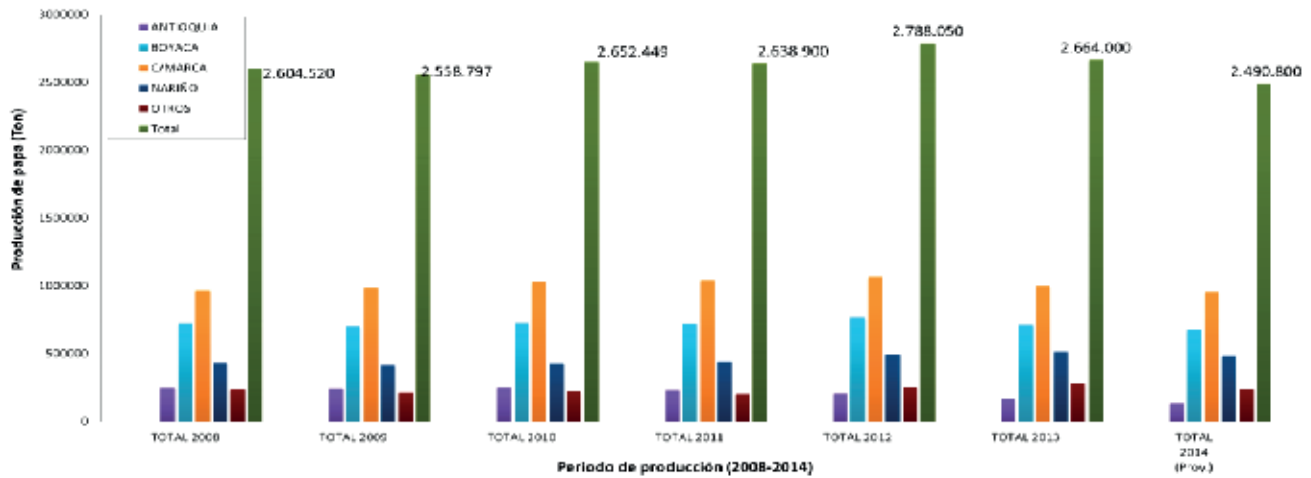


Fig. 1. Producción de papa en Colombia 2008-2014

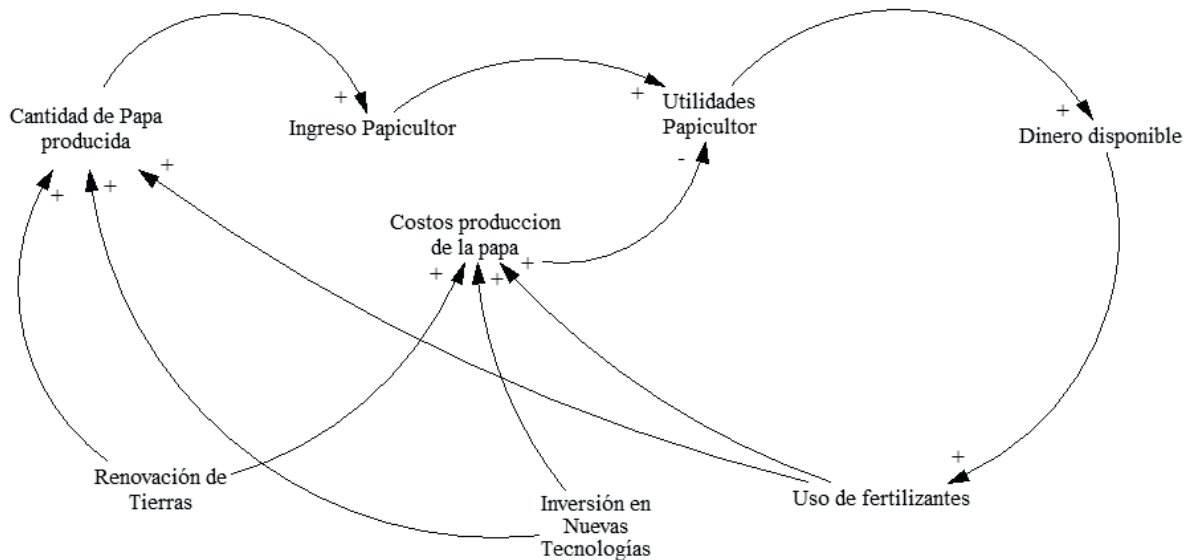


Fig. 2. Diagrama Causal

El diagrama causal de la Figura 2, tiene la siguiente lectura: A mayor cantidad de papa producida, el papicutor obtendrá más ingresos y, por ende, más utilidades. Si la cantidad de utilidades es buena, habrá mayor cantidad de capital disponible para invertir en dichas estrategias; de lo contrario, este no podrá realizarlas. La idea fundamental con este modelo es analizar el efecto que tiene la volatilidad del precio del tubérculo y el impacto que tiene la aplicación de dichas estrategias (uso de fertilizantes, renovación de tierras, inversión en nuevas tecnologías) sobre las utilidades del papicutor. En la Figura 2, se presenta el diagrama causal del modelo.

### 3.2. Diagrama de flujo y niveles

En el modelo se consideró el precio de la papa como una constante  $P = 1.300 \text{ \$ / kilo}$  (y se asumió que crece o decrece con la inflación. La inflación se consideró como una constante cuyo valor es de 4,5% y el número de semillas sembradas considerado fue de .10.000.

Considerando el efecto de la inversión en tecnología, la fertilización y la renovación de tierras, se asumió que la cantidad final producida corresponde al pro-

ducto de las semillas cultivadas por unas constantes cuyos valores representan los efectos de la implementación de dichas estrategias (1):

$$\text{Producción papa} = N^{\circ} \text{ de semillas plantadas} * (1.65) \quad (1)$$

Los ingresos del papicultor corresponden al producto entre el precio y la producción de papa (2):

$$\text{Ingresos papicultor} = \text{Precio} * \text{producción} \quad (2)$$

La inversión en nuevas tecnologías y el costo de la mano de obra se asumieron como constantes, con incrementos mensuales. El número de hectáreas sembradas considerado fue de 60 y el precio de los fertilizantes crece con un incremento constante (3), (4), (5) y (6).

$$\text{Inversión nuevas tecnologías} = 1.000.000 * \text{Inflación} \quad (3)$$

$$\text{Costo mano de obra} = 3000 \$ - \text{hectárea} * \text{incremento} \quad (4)$$

$$N^{\circ} \text{ hectáreas} = 60 \quad (5)$$

$$\text{Precio fertilizantes} = 560\$ / \text{hectáreas} * \text{incremento} \quad (6)$$

La renovación de tierras corresponde al producto entre el número de hectáreas y el costo de la mano de obra. Ver ecuación (7):

$$\text{Renovación Tierras} = \text{Costo mano de obra} * N^{\circ} \text{ de hectáreas} \quad (7)$$

Los costos de fertilización corresponden al producto entre el número de hectáreas y el precio de los fertilizantes. Ver ecuación (8):

$$C \text{ Fertilización} = N^{\circ} \text{ de hectáreas} * \text{Precio fertilizantes} \quad (8)$$

La suma de los costos derivados de la renovación de tierras y los costos de fertilización permiten calcular el costo total (9):

$$\text{Costos} = \text{Renovación de tierras} + \text{costos fertilización} \quad (9)$$

Por último, las utilidades corresponden a la diferencia entre los ingresos y los costos totales (10)

$$\text{Utilidad Neta} = \text{Ingresos} - \text{costos e inversiones} \quad (10)$$

En la Figura 3 se presenta el diagrama de Flujos y Niveles y en la Tabla 1 las variables y su clasificación según el tipo:

Tabla 1. Variables del modelo

Variable	Tipo
Precio de la papa	Nivel
Número de semillas	Parámetro
Producción de papa	Auxiliar
Ingresos papicultor	Auxiliar
Inversión en nuevas tecnologías	Parámetro
Costo mano de obra	Parámetro
Número de hectáreas	Parámetro
Precio de los fertilizantes	Parámetro
Renovación de tierras	Auxiliar
Costos de fertilización	Auxiliar
Costos e inversiones	Auxiliar
Utilidades netas	Auxiliar

### 3.3. Resultados obtenidos

En la Figura 4 y 5, se observa que la implementación de las estrategias de fertilización, renovación de tierras e inversión en tecnología, permite al papicultor aumentar sus utilidades considerablemente. Si la inflación se asume positiva, el precio de la papa se incrementa con el paso del tiempo y así se obtendrán más utilidades. Los costos e inversiones también se incrementan; sin embargo, el papicultor tendrá recursos para cubrirlos. Las utilidades netas se incrementan paulatinamente a medida que pasen los meses.

En la Figura 6 y 7, se observa que si la inflación es negativa, el precio de la papa disminuye constantemente y, en consecuencia, el papicultor obtendrá menos utilidades. Los costos e inversiones se incrementarán; sin embargo, este podrá cubrirlos a pesar de que su margen de utilidades netas se reduzca. Las utilidades netas se irán incrementando paulatinamente, pero no tan rápido como en el caso anterior, a medida que pasen los meses.

### 3.4. Validación del modelo

La validación del modelo se realizó comparando los resultados con estadísticas e información del DANE, MinAgricultura y FEDEPAPA. Durante el año 2014, uno de los productos que más resultó afectado por diversos factores asociados a la oferta fue la papa [8], lo cual llevó a que los agricultores protestaran por los bajos precios que les estaban pagando por su producto y por los altos costos de producción [8].

Según FEDEPAPA, las fuertes variaciones en los precios del tubérculo generadas por el fenómeno de El Niño y las afectaciones en ciertas zonas con po-

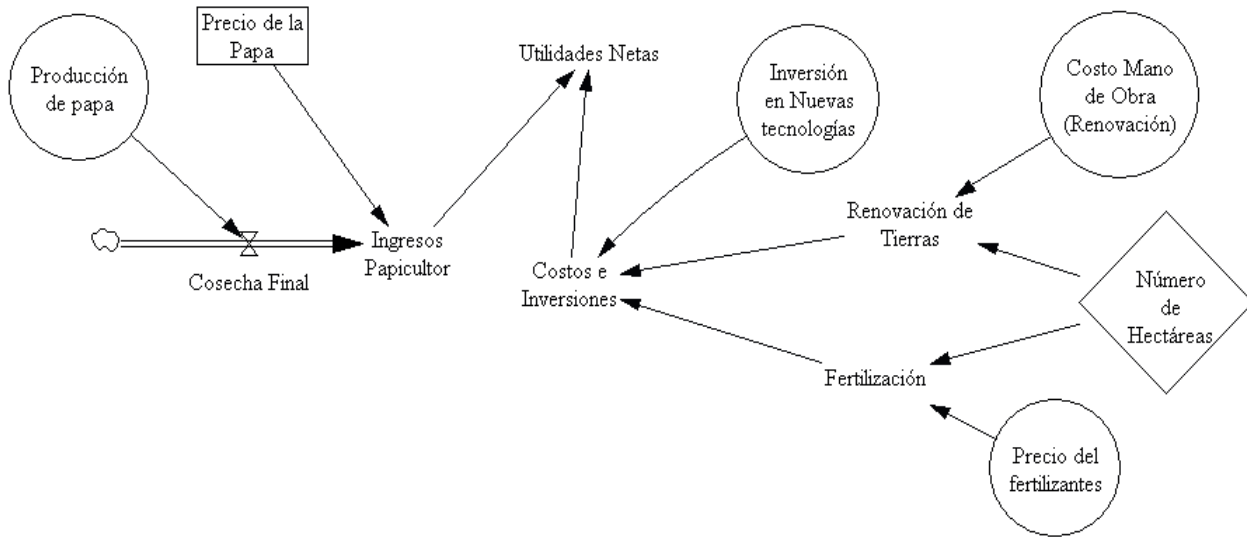


Fig. 3. Diagrama de Flujo y Niveles

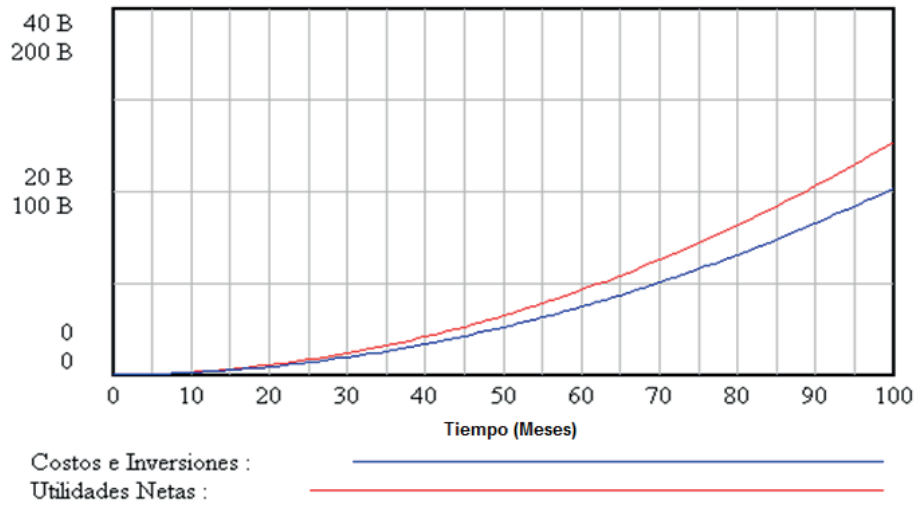


Fig. 4. Utilidades Vs Costos e Inversiones (en \$) (B= mil millones)

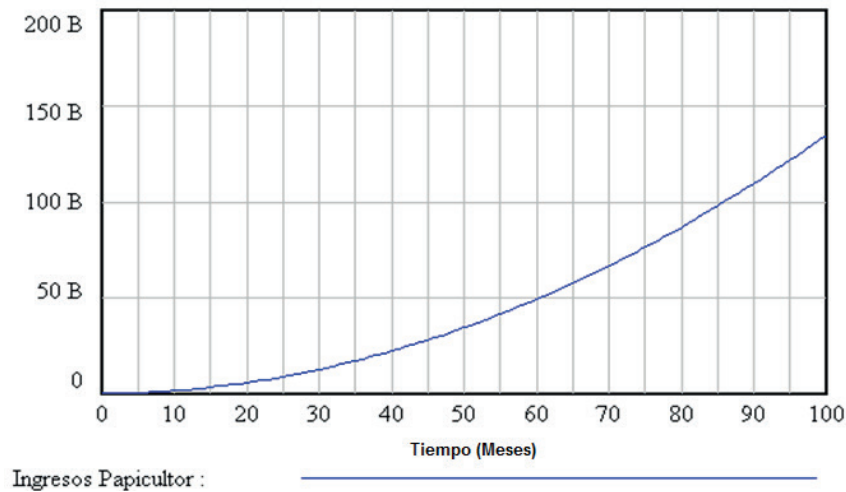


Fig. 5. Ingresos Papicultor (Inflación positiva) (en \$) (B= mil millones)

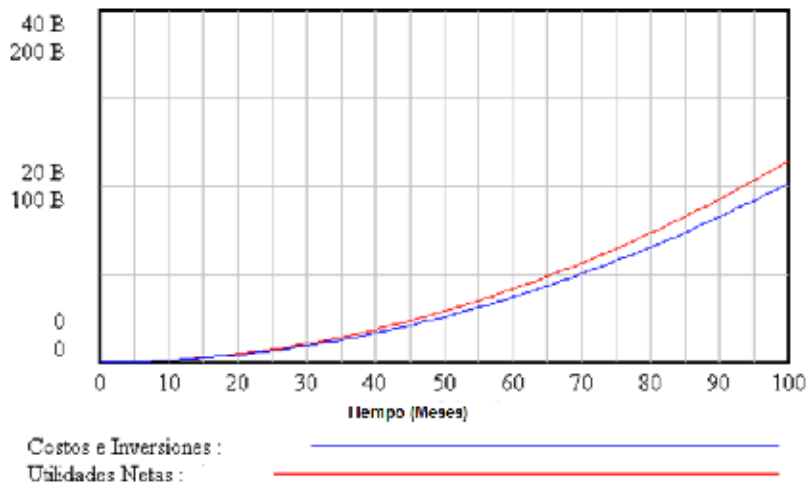


Fig. 6. Utilidades Vs Costos e Inversiones (en \$) (B= mil millones)

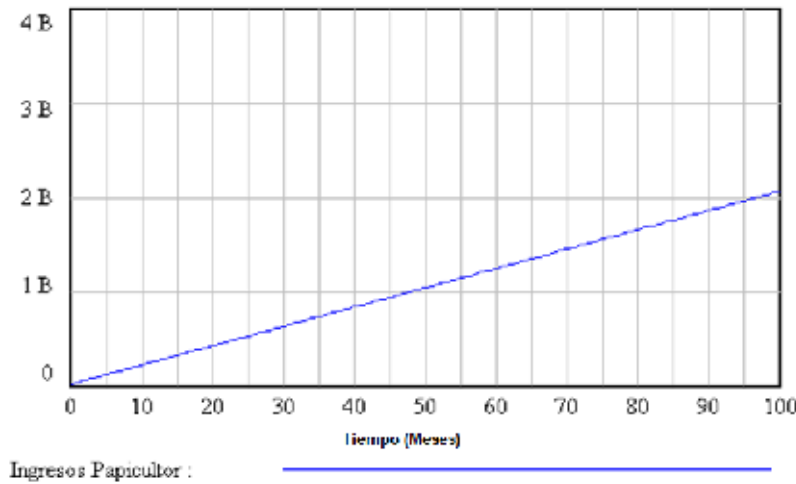


Fig. 7. Ingresos papicultor (Inflación negativa) (en \$) (B= mil millones)

blaciones de insecto-plagas, como la polilla guatemalteca, fueron dos factores nocivos que dificultaron la producción del tubérculo durante ese año [9]. Ante esto, los papicultores se vieron obligados a adoptar medidas preventivas (estrategias de fumigación y fertilización) para no impactar notablemente la inflación, lo que valida los supuestos y resultados del modelo [9].

En los períodos de mayor aporte, el 30% de la inflación puede ser explicada por los cambios en los precios de la papa [3]. Su ponderación en la inflación, medida por el índice de precios al consumidor (IPC) base 88, era de 1.09, siendo después de la carne de res, la carne de pollo y el arroz, el alimento más importante [3]. Disminuyó a 0.98 en el IPC base 98,

debido posiblemente a los cambios en la canasta y a la introducción de nuevos bienes. Aun así, la papa continúa teniendo una ponderación importante en relación con el resto de los alimentos [3].

### 3.5. Políticas que ayudarían a mejorar las condiciones del problema

Algunas políticas que mejorarían las condiciones del problema estarían relacionadas con la estabilización de los costos de producción y la diversificación. Para estabilizar la producción de papa y aumentar la producción a través del tiempo, se sugiere realizar inversiones en tecnología cada 2 o 3 años y renovar las tierras cada 3 años.

La diversificación con cultivos complementarios a la papa, no solo ayudan a aumentar los ingresos, sino que también reducen la dependencia de los papicultores con el precio de la papa. La diversificación también ayuda a que ellos puedan atravesar por temporadas más largas de precios bajos de la papa, sin verse tan afectados económicamente, pues, con la diversificación, pueden comercializar con otra clase de tubérculos.

Finalmente, la estrategia que prioriza la diversificación con respecto a las otras estrategias parece ser la que genera mejores resultados, adoptando diferentes escenarios climáticos y de precios.

#### 4. CONCLUSIONES

Las utilidades del papicultor colombiano son especialmente sensibles a tres factores: precio, costos de producción e inflación. La papa es un producto básico que no es objeto de intercambios mundiales y sus precios dependen generalmente de la oferta y la demanda locales. El Gobierno colombiano debe otorgar subsidios a los papicultores para que tengan recursos que les permitan realizar inversiones en tecnología, incrementar la producción y las utilidades. Los papicultores colombianos tienen un amplio conocimiento sobre el cultivo, consumo y comercialización de la papa, siendo un cultivo profundamente arraigado en la lógica del pequeño agricultor colombiano, razón por la cual se deben crear y establecer políticas que favorezcan su producción.

Si los papicultores adoptan un pensamiento sistémico, pueden establecer estrategias de diversificación de la producción. En épocas en las cuales el precio de la papa es bajo o las condiciones climáticas no sean las mejores, estos pueden optar por cosechar otra clase de tubérculos.

Se pueden desarrollar otros modelos de simulación en otros aspectos relacionados con la producción de papa en Colombia u otra clase de tubérculos que tienen una representación significativa en la economía del país.

#### REFERENCIAS

- [1] DINERO, “Minagricultura radica proyectos de ley para crear fondos parafiscales de la papa y el fique,” 2012. Disponible en: <http://www.dinero.com/imprimir/164416>. [Consultado: Abril 21, 2015].
- [2] G. Farfán Pinzón, Milena, Peña Ardila, Geimar, “Evaluación de políticas de protección en el sector papicultor mediante sistemas dinámicos: Caso Villapinzón (Cundinamarca)”, Ingeniería, *Universidad Distrital Francisco José de Caldas*, vol. 20, no. 1, pp. 104-106, 2015.
- [3] C. Espinal, H. Martínez, N. Pinzón, and C. Barrios, *Ministerio de Agricultura y Desarrollo Rural: Observatorio Agrociudades Colombia, Documento de Trabajo No. 54,* Bogotá, pp. 5-8, 2005.
- [4] C.D. Cantillo Bertel, A.M. Patiño Quiróz, K. Bermúdez Salcedo, K. Palma de Ávila, P.A. De la Hoz, *Análisis del comportamiento de la producción y exportación de papa en Colombia entre el periodo 2008-2014*, pp. 22-30, 2014.
- [5] N. Estrada y L. Valencia, “Desarrollo de cultivares de papa resistentes a la palomilla,” *Revista Latinoamericana: La Papa*, vol. 1, pp. 64-73, 1988.
- [6] MINAGRICULTURA, *MinAgricultura anuncia modernización y promoción del consumo para el sector papicultor*, 2014. .
- [7] F. A. Osorio, S. A. Aramburo, and Y. O. Morales, “Simulación de estrategias de inversión para pequeños papicultores colombianos. Simulation of Investment Strategies for Small Coffee Farmers in Colombia,” *Revista Facultad Nacional de Agronomía*, vol. 64, no. 65, pp. 6277-6290, 2011.
- [8] MINAGRICULTURA, “Boletín semanal Precios mayoristas MINAGRICULTURA 2015,” *SIPSA*, 2015. Disponible en: [https://www.dane.gov.co/files/investigaciones/agropecuario/sipsa/Sem\\_03ene\\_09ene\\_2015.pdf](https://www.dane.gov.co/files/investigaciones/agropecuario/sipsa/Sem_03ene_09ene_2015.pdf).
- [9] Colprensa, “‘El niño’ desata alerta entre papicultores”, Colprensa 2014. Disponible en: <http://m.la-tarde.com/actualidad/colombia/134930-el-nino-desata-alerta-entre-papicultores>.



# Un Caso de Estudio para la Adopción de un Modelo de Trazabilidad de Requisitos en el Sector Energético

## A Case Study for the Adoption of a Requirement Tracking Model for Energy Sectors

**Lina María Montoya-Suárez, MSc**

*Facultad de Ingenierías y Arquitectura  
Fundación Universitaria Luis Amigó  
Medellín, Colombia.  
lina.montoyasu@amigo.edu.co*

**Juan Carlos Monsalve-Gómez, MSc**

*Facultad de Ingenierías y Arquitectura  
Fundación Universitaria Luis Amigó.  
Medellín-Colombia  
juan.monsalvego@amigo.edu.co*

**Jorge Mauricio Sepúlveda-Castaño, MSc**

*Facultad de Ciencias Básicas e Ingeniería  
Corporación Universitaria Remington  
Medellín, Colombia,  
jsepulveda@remington.edu.co*

(Recibido el 09-09-2014. Aprobado el 12-12-2014)

**Resumen.** Es fundamental para un desarrollo de software, mejorar el proceso relacionado con los requisitos, los aspectos que facilitan una mejora, es contar con la posibilidad de realizar trazados a los requisitos a lo largo del ciclo de vida de un sistema. Esta trazabilidad permite evaluar las implicaciones que puede tener el cambio y posibilita su detección en fases tempranas o incluso en la fase de gestión de la configuración, siguiendo las huellas de trazado.

A través de este artículo de investigación, se presenta un caso de estudio aplicando la herramienta Enterprise Architect para el sector energético, en el que se evidencia cómo se puede dar seguimiento a un requisito, comenzando en la primera fase de análisis y diseño (requisitos funcionales, requisitos no funcionales, regla de negocio, casos de uso, diagrama de clase, diagrama de componente, etc.), por medio del cual es posible generar una matriz de trazado dentro de un proceso de desarrollo de software.

**Palabras claves:** Trazabilidad, trazabilidad de requisitos, arquitectura empresarial.

**Abstract.** In order to improve software development process relating to requirements for improvement, it is necessary to have a possibility of requirements drawn along the life cycle of a system. For this reason, traceability allows evaluating the potential impacts of change so we can detect changes at an early stage or even at the stage of configuration management in the footsteps of software development.

Through this paper a case study is presented using the Enterprise Architect tool applied to the energy sector while evidencing how to track a requirement beginning in the first phase of analysis and design (functional requirements, non-functional requirements, business rules, use cases, class diagram, component diagram etc.) generating array layout within a software development process.

**Keywords:** Traceability, requirements traceability, enterprise architecture.

**Citación de artículo, estilo IEEE:**

L.M. Montoya-Suárez, J.C. Monsalve-Gómez, J.M. Sepúlveda-Castaño "Un Caso de Estudio para la Adopción de un Modelo de Trazabilidad de Requisitos en el Sector Energético", *Lámpsakos*, Nº 13, pp. 88-100, 2015.

DOI: <http://dx.doi.org/10.21501/21454086.1494>



## 1. INTRODUCCIÓN

En la Ingeniería de Software, hacer trazabilidad en los requisitos permite obtener en el producto soluciones más exactas y fiables, con base en las necesidades solicitadas por el cliente. La trazabilidad adquiere gran importancia debido a que es la “aptitud para rastrear la historia, la aplicación o la localización de una entidad mediante indicaciones registradas” [1][2].

Es fundamental hacer el seguimiento a los requisitos cuando se desea construir un software de calidad, lo cual se constituye en un proceso de análisis y estudio detallado. Cuando ocurren cambios en el software, la trazabilidad hace que sea relativamente más fácil evaluar el impacto que los cambios podrían tener en otras partes del proceso de desarrollo; esto permite seguir las huellas en los requisitos [3].

Es primordial para la fase de desarrollo de software poder fijar líneas bases desde el comienzo hasta el final de las etapas, logrando una perspectiva horizontal; una de la mejores prácticas consiste en rastrear las actividades mismas, las historias que estas presentan, pues permiten a los desarrolladores no caer en los mismos errores, buscando soluciones en menos tiempo y mayor rendimiento. Para describir y seguir los requisitos, se genera en dos direcciones: hacia delante (forward) y hacia atrás (backward), como lo expresa Ramesh y Jarke [4].

Los errores en la gestión de requisitos pueden ser muy costosos para las empresas; contar con un mecanismo integral de trazabilidad facilita la finalización de las tareas relacionadas con pruebas y documentación, para esto la trazabilidad garantiza que todos los requisitos sean diseñados y que todos los diseños se codifiquen y se prueben, aunque no solo resalta las pruebas que se deben actualizar o repetir, sino que también señala los documentos (análisis de riesgos, especificaciones y manuales del usuario) [5].

El objetivo fundamental de este artículo de investigación es, mediante el proyecto IDICA, presentar un caso de estudio para el sector energético, con el propósito de hacer matrices de trazado de los requisitos y así rastrear las historias, haciendo seguimiento en diferentes dimensiones de análisis y diseño como lo permite la herramienta Enterprise Architect. Esta contribución técnica se estructura de la siguiente manera: En la sección 2, se presentan las generalidades de trazabilidad de requisitos. En la sección 3, se

estructuran los elementos y metodología empleada en la investigación. En la sección 4, los resultados obtenidos de las matrices generadas y, finalmente, se presentan las conclusiones.

## 2. GENERALIDADES DE TRAZABILIDAD DE REQUISITOS

Cuando un requisito pasa por un proceso de calidad es necesario tener la confianza acerca de la corrección y viabilidad del requisito por el cual permite hacer preguntas como [6]: ¿Qué ocurre con la especificación de los requisitos en conjunto?

Para la especificación de los requisitos, se hace referencia a la colección de requisitos especificados y definidos previamente. La especificación no tiene que estar en un determinado formato, puede ser una especificación sobre papel, o un blog, o algo similar [6][3][7].

Cuando la especificación de los requisitos está completa, se tendrá un conocimiento preciso del alcance, el impacto, la funcionalidad del producto y el costo asociado. Este es el momento de llevar a cabo la revisión de la especificación, en la cual se validan todos los requisitos.

Es importante tener en cuenta y asegurarse de que los requisitos tengan consistencia y completitud; en caso contrario, cualquier conflicto entre los requisitos ha sido resuelto [6]. Dos requisitos están en conflicto si no pueden implementarse juntos, es decir, si la solución a un requisito impide la implementación de otro.

Hay que observar cuando las expectativas del cliente son altas, es elemental tener en cuenta la fecha de entrega y la disponibilidad de los recursos; además, asegurarse de que las funciones más importantes del producto sean entregadas tan pronto como sea posible. Otro problema que puede suceder es que haya demasiados requisitos. La solución a los problemas anteriores es la priorización de los requisitos funcionales, no funcionales y la de regla de negocio [6] [8].

A continuación se puede observar cómo se clasifican los requisitos:

**Tabla 1.** Clasificación de los requisitos de software [9].

Factor de Calidad	Atributos
Funcionalidad	Característica y capacidades del Programa
	Generalidad de las funciones
	Seguridad del sistema
Facilidad de uso	Factores humanos
	Factores estéticos
	Consistencia de la interfaz
	Documentación
Confiabilidad	Frecuencia y severidad de los fallos
	Exactitud de las salidas
	Tiempo medio de fallos
	Capacidad de recuperación ante fallos
	Capacidad de predicción
Rendimiento	Velocidad de procesamiento
	Tiempo de respuesta
	Consumo de recursos
	Rendimiento
	Eficacia
Capacidad de Soporte	Extensibilidad
	Adaptabilidad
	Capacidad de prueba
	Capacidad de configuración
	Compatibilidad
	Requisitos de instalación

### Matriz de Trazabilidad

Un concepto clave en el proceso de gestión de cambios es la trazabilidad. Los requisitos deben ser trazables, es decir, “rastreados”. Se podría decir que un requisito es trazable si se pueden identificar todas las partes del producto existente relacionadas con ese requisito. Todos los requisitos deberían ser trazables para mantener consistencia entre los distintos documentos de un proyecto [7].

Es importante conocer aspectos de los requisitos tales como [8]:

- Su origen (Quién los propuso).
- Necesidad (Por qué existe).
- Relación con otros requisitos (Dependencias).
- Relación con otros elementos (Dependencias).

Una de las claves para la gestión de cambio es la matriz de trazabilidad, la cual permite establecer las relaciones entre los diferentes requisitos desde un

punto de vista jerárquico y la documentación generada durante todo este proceso es la base sobre la que debe construirse todo el plan de pruebas. El uso de matrices de trazabilidad es una buena técnica para llevar a cabo esta actividad de forma eficiente [9] [10].

En la matriz se irán registrando los requisitos de negocio. Por cada requisito de negocio, se identificarán los requisitos de usuario correspondientes. De cada requisito de usuario se identificarán cuáles son los requisitos de sistema asociados a cada uno de ellos. Y así sucesivamente, se irá llenando toda la matriz de requisitos con el fin de gestionar [11].

La siguiente matriz se utiliza para relacionar requisitos. Es una matriz de dependencias, como se muestra en la Figura 1.

En este caso, los requisitos X representan los requisitos que originan las dependencias y los requisitos Y serían los requisitos que dependen de otros requisitos, de los requisitos X.

Por ejemplo, se puede ver cómo los requisitos Y 1,4 y 7 dependen del requisito X 2, o también depende del requisito X 5. También se observa cómo los requisitos Y 2 y 5 dependen del requisito 1. De esta forma, se puede ver de qué manera se relacionan los requisitos, para analizar mejor el impacto de los cambios.

Otro tipo de matriz de dependencias muy usada es la *matriz hacia atrás / hacia adelante*

- Requisitos vs especificación
- Especificación vs caso de prueba
- Especificación vs diseño
- Diseño vs código

La construcción de estas matrices trae beneficios más allá de un simple registro de la correlación o dependencia entre los elementos de los modelos [11]. A partir de ellas, es posible analizar o inferir información acerca del nivel de especificación de los requisitos, el nivel de participación de los usuarios, el costo asociado a cada fase de desarrollo, la arquitectura requerida, el plan de las pruebas, etc. Si las matrices son actualizadas continuamente, pueden usarse como herramienta de soporte para el personal encargado del mantenimiento y las pruebas del Sistema [12].

		REQUISITOS (X)					
		R1	R2	R3	R4	R5	Rn
REQUISITOS (Y)	R1		X				X
	R2					X	
	R3						
	R4		X	X			X
	R5					X	
	R6						
	R7		X	X			X
	Rn			X		X	

Fig. 1. Matriz de dependencias requisitos X vs requisitos Y [6] [1].

### 3. ELEMENTOS Y METODOLOGÍA

Para el caso de adopción de un modelo de trazado, para el grupo de desarrollo INDICA (Índices de Calidad), se generó una matriz de trazabilidad para las necesidades, con el fin de garantizar la consistencia y la completitud de los modelos de desarrollo. Mediante estudio de casos, según Fong, enmarcada dentro de cinco fases, se definió lo siguiente [13]:

#### A. El fenómeno que se aborda en el estudio

El objetivo es aplicar un caso de estudio por medio de la herramienta Enterprise Architect, haciendo matrices de trazado para el proyecto INDICA, en el que se evidencia cómo se puede dar seguimiento al levantamiento de los requisitos.

#### B. Proposición

La investigación que se expone pretende responder la siguiente pregunta: ¿Es posible generar seguimiento de los requisitos por medio de modelo de trazado para el proyecto INDICA?

#### C. Unidades de análisis de la investigación

Estas unidades de análisis son:

- 1) Identificar los modelos de trazado y su aplicabilidad.
- 2) Aplicación del proyecto INDICA por medio de un modelo de trazado para el levantamiento de los requisitos.
- 3) Generación de matrices de trazabilidad.

#### D. Definición de lógica que relaciona los datos con las hipótesis del estudio

- 1) Trazabilidad de requisitos, desarrollo de software, calidad en el desarrollo de software.
- 2) A partir de la teoría de la trazabilidad de requisito, identificar cuál es la evidencia necesaria para verificar y garantizar la objetividad.
- 3) De un caso de estudio, evidenciar el modelo de trazado con base al levantamiento de requisitos.

#### E. Criterios para interpretación de datos obtenidos

- 1) Matriz de trazabilidad
- 2) Necesidades vs stakeholders
- 3) Necesidades vs requisitos
- 4) Necesidades vs proceso
- 5) Casos de uso vs validaciones y mensajes del sistema
- 6) Casos de uso vs regla de negocio
- 7) Casos de uso vs requisitos funcionales generales
- 8) Casos de uso vs requisitos no funcionales
- 9) Casos de uso vs diagrama de actividades
- 10) Requisito no funcional vs requisitos funcionales generales
- 11) Requisitos vs interfaces.

El proyecto se trata de un sistema INDICA que permite cargar datos de los operadores de red del sistema eléctrico nacional, acerca de las indisponibilidades de sus circuitos y transformadores que generar índices, con el fin de realizar auditoría y control de cali-

dad, lo cual redundará en beneficio de los operadores de distribución local.; este sistema genera proceso de calidad para los operadores de distribución local.

El modelo de trazabilidad propuesto tiene el objetivo de orientar el uso de la matriz de trazabilidad de los requisitos, casos de usos y necesidades del sistema. Está compuesto por un conjunto de pasos que se explican por medio del desarrollo del caso de estudio INDICA para calcular índices del sistema de distribución local.

Se usó para el caso de estudio la herramienta Enterprise Architect. Es una plataforma de modelado, diseño y administración, colaborativa, basada en UML 2.3 y estándares relacionados. Ágil, intuitiva y extensible, permite visualizar, analizar, modelar, probar y mantener un amplio rango de sistemas, software, procesos y arquitecturas [14]

Esta herramienta provee trazabilidad completa desde los modelos de requisitos, análisis y diseño, hasta la implementación y despliegue. La verificación y validación efectiva y el análisis del impacto inmediato son posibles a través del ciclo de vida completo, usando capacidades tales como la matriz de relaciones y la vista de jerarquía de Enterprise Architect [15].

#### 4. RESULTADOS

Actualmente, se mide la calidad del servicio que presta el Sistema de Distribución Local (SDL) en Medellín-Colombia, para que el usuario final perciba los beneficios o perjuicios relacionados con este y el Operador de Red (OR) tenga un incentivo para prestar el mismo con ciertos estándares de eficiencia.

A raíz del inicio de un nuevo periodo tarifario, cambia la metodología para medir la calidad del SDL (resolución CREG (Comisión Reguladora de Energía y Gas 097/2008), por lo que es necesario definir un sistema de información para respaldar este hecho. Expertos en Mercados (XM), en su calidad de LAC (área de liquidación de cuentas en XM), se ve impactado debido a que se le asignan nuevas responsabilidades, de tal manera que estará encargado de gestionar las actividades relacionadas con el cálculo de índices de calidad en el SDL. Adicionalmente los OR se ven afectados, ya que dependiendo del comportamiento

de los índices, podrán tener posibles variaciones en sus ingresos (a favor o en contra), vía cargos por uso del SDL.

La Figura 2 muestra los *stakeholders* directos e indirectos. Para los directos, presenta 9 actores: La entidad reguladora tiene asociados los actores SSPD (Administrador del sistema SUI, a través del cual se obtiene la información relacionada con las interrupciones declaradas por los OR y la información del PAR Programación de activo en reparación) y CREG (Encargado de definir las reglas de negocio que deben ser usadas para llevar a cabo el cálculo de índices de calidad del SD). Para los indirectos, 11 actores, de los cuales el líder técnico tiene asociados a 4.

##### 4.1. Matriz de trazabilidad

Al tener toda la información recopilada en artefactos, en Enterprise Architect se obtienen las matrices de trazabilidad que se relacionen a cualquier elemento del sistema INDICA. Las matrices de trazabilidad y las restricciones ayudan al proyecto a verificar que la metodología se cumpla correctamente [15].

Las matrices de trazabilidad tienen una doble funcionalidad:

- Obtener las plantillas por cada tipo de artefacto que hacen referencia a otros elementos relacionados.
- Permitir la trazabilidad de los elementos entre las fases.

Para el proyecto INDICA se han definido las siguientes matrices de trazabilidad:

- 1) Necesidades vs. *stakeholders*
- 2) Necesidades vs. requisitos
- 3) Necesidades vs. procesos
- 4) Casos de uso vs. validaciones y mensajes del sistema
- 5) Casos de uso vs. regla de negocio
- 6) Casos de uso vs. requisitos funcionales
- 7) Casos de uso vs. requisitos no funcionales
- 8) Casos de uso vs. diagrama de actividades
- 9) Requisito no funcional vs. requisitos funcionales generales

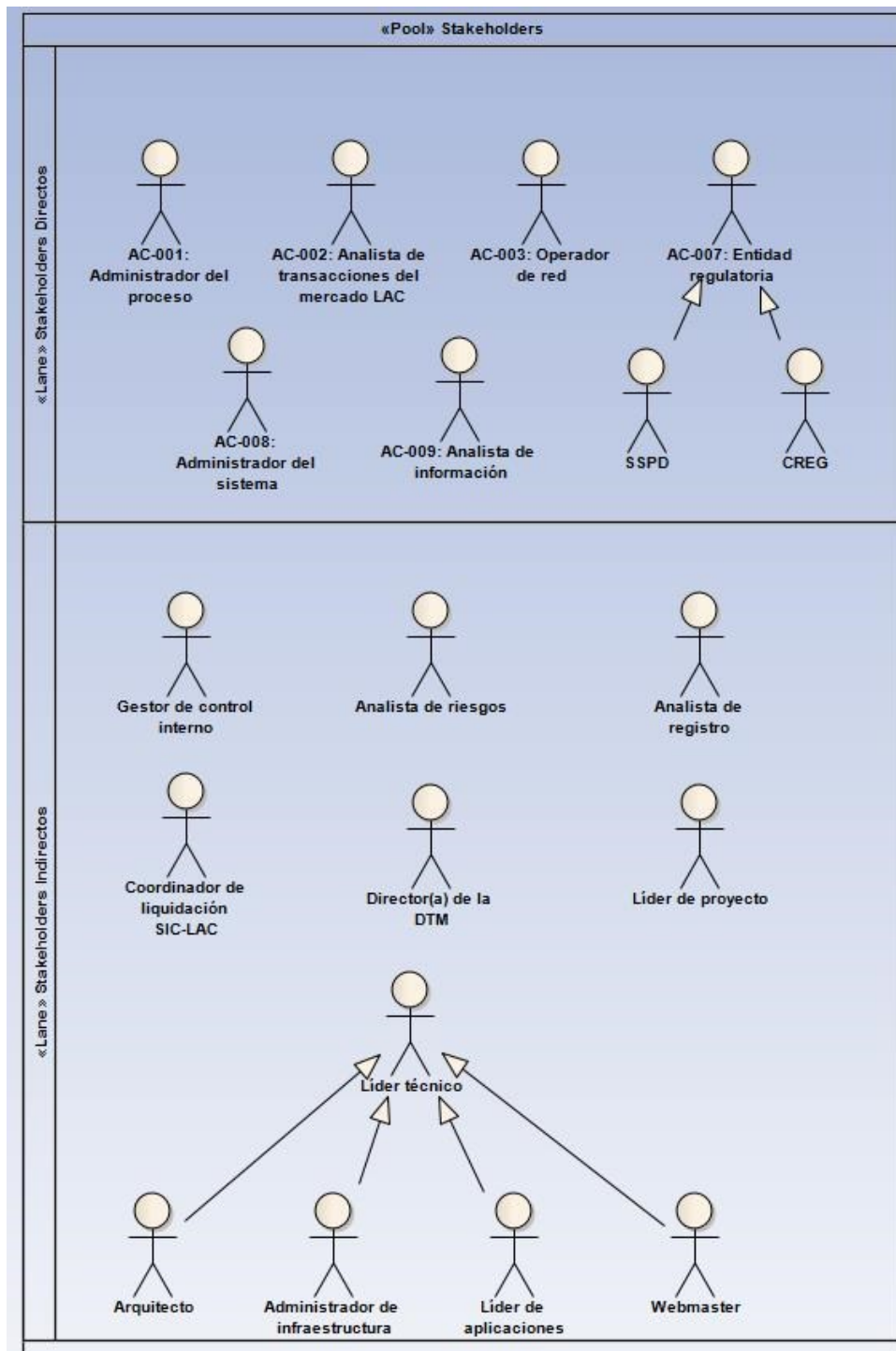


Fig. 2. Stakeholders directos e indirectos [7].

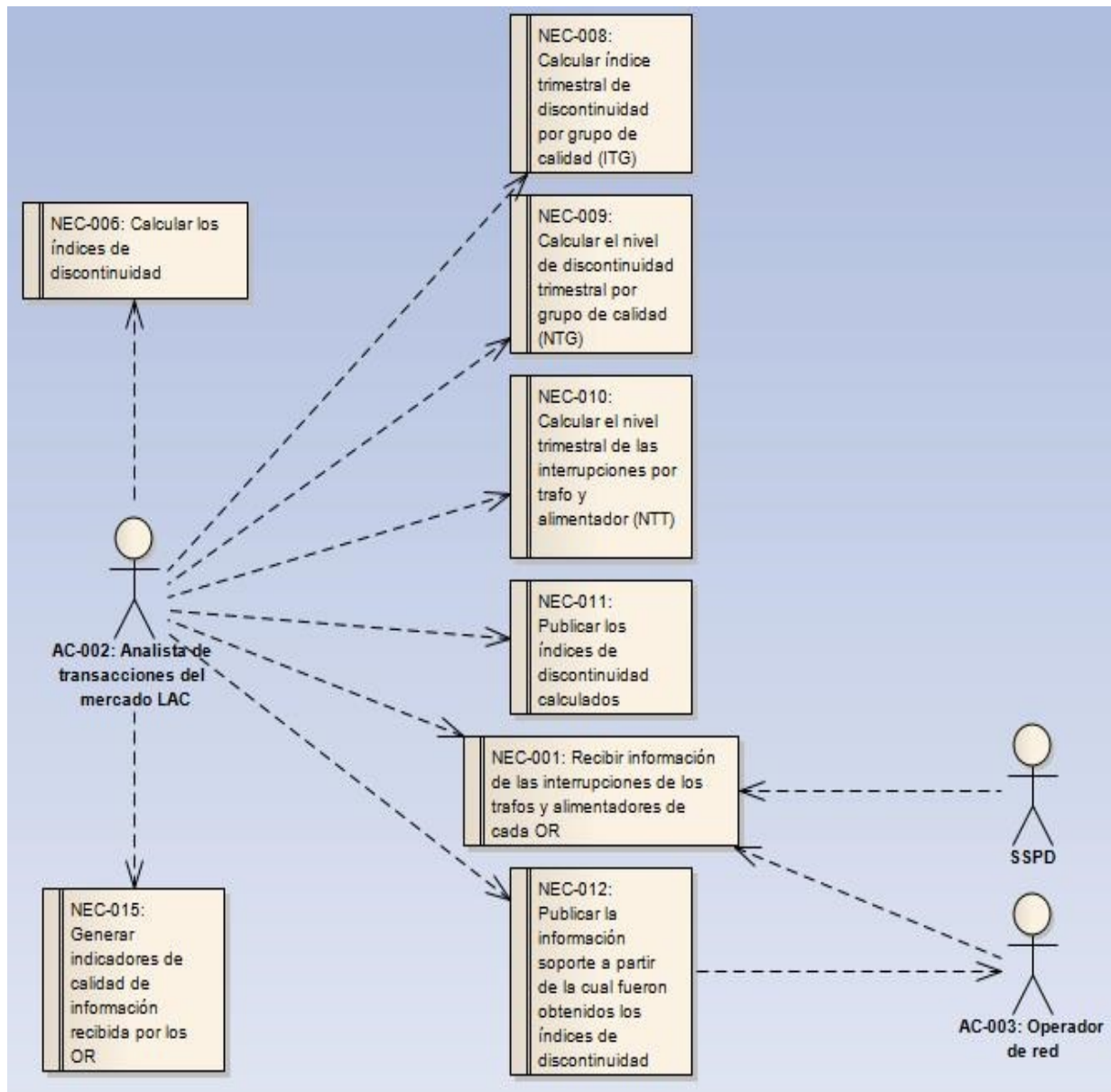


Fig. 3. Necesidades vs. stakeholders [7].

#### 10) Requisitos vs. interfaces generales

Las matrices de trazabilidad se generan y se consultan accediendo a la vista de recursos (resources) del Enterprise Architect.

#### 4.2. Selección de matriz de trazabilidad

Una vez que el proceso de desarrollo de software es seleccionado por el equipo de desarrollo, ellos crean o seleccionan matrices de trazabilidad después de atender las sugerencias definidas.

Aunque el patrón facilite la creación de diferentes tipos de modelos de trazado para soportar las actividades de gestión de la configuración, el método de trazabilidad es probado con dos definiciones genéricas:

- 1) El de matriz de trazabilidad basado en casos de uso vs. clase. (Anexo 1)
- 2) El de matriz de trazabilidad basado en casos de uso vs. requisitos. (Anexo 2)

3) El de matriz de trazabilidad basado en casos de uso vs. sistema. (Anexo 3)

Para el método de matriz de trazabilidad basado en necesidades, el vínculo de trazado se define de la siguiente forma:

4) Necesidades asociado con los stakeholders. (Figura 3)

La selección de la matriz de trazabilidad para el alcance del sistema permite dar solución al flujo de información requerido en empresas del sector energético, pues se pueden gestionar los cálculos de los índices de calidad del SDL y hacerlos públicos a los stakeholders involucrados directamente, es decir que se facilita el análisis y comportamiento de los índices.

El lector puede referirse a los siguientes anexos:

**Anexo 1:** Matriz de trazabilidad casos de uso vs. requisitos funcionales: De esta forma, se puede ver de qué manera se relacionan los casos de usos con los requisitos funcionales y así analizar mejor el impacto de los cambios.

**Anexo 2:** Matriz de trazabilidad casos de uso vs clase: De esta forma, se puede ver de qué manera se relacionan los casos de usos con la clase y así analizar mejor el impacto de los cambios.

**Anexo 3:** La matriz de trazabilidad casos de uso vs sistema: De esta forma, se puede ver de qué manera se relacionan los casos de usos con los sistemas.

## 5. CONCLUSIONES

La presentación de este tema mediante un caso de estudio para el Proyecto de Software INDICA permitió generar matrices de trazabilidad, con la finalidad de que los desarrolladores, así como los ingenieros de requisitos, lograran tener un conocimiento pleno de los artefactos del sistema y su relación entre ellos, tanto dentro de la aplicación como con otras aplicaciones de software enlazados al proyecto, para la práctica de la trazabilidad y la transformación de modelos en cualquier entorno de desarrollo de software.

El marco conceptual, en primera instancia, estandariza el uso de los vínculos de trazado definidos en la herramienta Enterprise Architect [15], con el fin de controlar la transformación de modelos. Esto ayuda a eliminar una de las grandes incertidumbres de los desarrolladores, al usar una relación de dependencia para la trazabilidad, ya que sus definiciones son tan similares que no es fácil identificar la situación real de uso, además de la utilidad de usar los diferentes tipos ofrecidos por la herramienta.

La matriz de trazabilidad generada de manera horizontal y vertical permite un mejor cubrimiento de los artefactos involucrados en un sistema, de forma que se pueda hacer un seguimiento de trazado hacia adelante y hacia atrás dentro del desarrollo de software.

La práctica de la trazabilidad aplicada por la herramienta Enterprise Architect permite hacer una administración de los requisitos a la hora de generar un desarrollo de software. Es de suma importancia considerar el personal de desarrollo o calidad dedicado a la validación y verificación de la trazabilidad de requisitos, para que establezcan un modelo de trazado con el cual se pueda medir el nivel de exactitud de los modelos de desarrollo con respecto a los requisitos planteados por los usuarios y los clientes.

## REFERENCIAS

- [1] ISO for Standardization, ISO 8402: 1994: *Quality Management and Quality Assurance Vocabulary*. International Organization for Standardization, 1994.
- [2] B. G. Dale, T. Van Der Wiele, and J. Van Iwaarden, *Managing quality*, John Wiley & Sons, 640p., 2013.
- [3] V. Anaya and P. Letelier, "SmartTrace: Una herramienta para trazabilidad de requisitos en proyectos basados en UML", en *WER*, pp. 210-224, 2002.
- [4] B. Ramesh and M. Jarke, "Toward reference models for requirements traceability," *Softw. Eng. IEEE Trans.*, vol. 27, no. 1, pp. 58-93, 2001.

- [5] P. Letelier and V. Anaya, "Integrando especificaciones textuales y elementos de modelado UML en un marco de trabajo para trazabilidad de requisitos," en *JISBD*, pp. 2002, 151-162.
- [6] INTECO, Instituto Nacional de Tecnologías de la Comunicación, "Guía práctica de gestión de requisitos", España, 2008.
- [7] L. M. Montoya, "Levantamiento de un estado del arte para la trazabilidad de requisitos," Tesis de maestría, Universidad de Medellín, 2011.
- [8] J. A. Aguilar, I. Garrigós, and J. N. Mazón, "Modelos de weaving para trazabilidad de requisitos web en A-OOH," Actas del VII Taller sobre Desarrollo de Software *Dirigido por Modelos, JISBD, Congreso Español de Informática (CEDI), Valencia, España, SISTEDES*, 2010, pp. 146-155.
- [9] M. A. Carvajal, M.M. Sandoval, *La trazabilidad en el proceso de requerimientos de software*, 2008, 6p. Disponible en: <http://www.iiis.org/cds2008/cd2008csc/cisci2008/paperspdf/c601uz.pdf>
- [10] D. G. Firesmith, *Object-oriented requirements analysis and logical design: A software engineering approach*, John Wiley, New York, NY, USA: John Wiley & Sons, Inc., 1993, pp. 1-250.
- [11] J. Mendoza Castelán, *Unified Modeling Language*, Universidad Autónoma del Estado de Hidalgo, México, 2011, pp. 1-19.
- [12] B. W. Boehm, "Verifying and validating software requirements and design specifications," in *IEEE Software*, 1984, pp. 1-20.
- [13] C. F. Reynoso, "El estudio de casos en la investigación relacionada con la Pyme en México," *Revista Internacional de Investigación y Aplicación del Método de Casos*, vol. 21, p. 4, 2009.
- [14] J. W. Ross, P. Weill, and D. C. Robertson, *Enterprise architecture as strategy: Creating a foundation for business execution*, USA: Harvard Business Press, 2006, pp. 1-241.
- [15] E. Architect, *Sparx Systems*, 2012. Disponible en: <http://www.sparxsystems.com.au/>







Anexo 2: Matriz de Trazabilidad Casos de Uso vs Clase.

	Domain::AdministracionPlazoReporte	Domain::AgenteCalidad	Domain::Bitacora	Domain::BitacoraDetalle	Domain::Carga	Domain::CargaDiaria	Domain::CargaTrimestral	Domain::CausaEvento	Domain::Documento	Domain::GrupoPlazo	Domain::GrupoPlazoAgente	Domain::Hoja	Domain::ProcesoBitacora	Exception::Res097SDLException	Negocio::AdministracionService	Negocio::CargaEventosService
Administración::CU-011: Administrar perfiles del sistema		X												X		
Administración::CU-012: Asignar perfiles a usuarios del sistema					X			X				X				
Administración::CU-025: Autorizar OR para SDL												X				
Administración::CU-026: Definir plazos máximos						X									X	
Administración::CU-027: Definir plazos máximos de carga de información															X	
Administración::CU-028: Definir plazos máximos de cálculos y publicación	X					X	X	X								
Administración::CU-029: Administrar los tipos y códigos de los tipos de interrupción			X							X				X		
Administración::CU-030: Definir rangos de validez de datos								X								
Administración::CU-031: Administrar documentos					X						X				X	
Administración::CU-032: Administrar hojas	X				X										X	
Administración::CU-033: Autorizar recarga de información a los OR						X		X							X	
Administración::CU-042: Cambiar contraseña (Eliminado)								X								
Carga de Información::Administrador mapeo de OR y comercializador			X						X							
Carga de Información::CU-001: Descargar información del PAR			X													
Carga de Información::CU-004: Cargar información mensual OR desde un archivo de datos					X											X
Consultas y Reportes::CU-040: Consultar indicadores	X							X								X
Consultas y Reportes::CU-041: Consultar bitácora del sistema																
Consultas y Reportes::CU-056: Generar reporte consolidado por elemento												X				
Procesar Información::CU-024: Publicar índices e insumos								X								
Procesar Información::CU-050: Verificar factibilidad de ejecución de los cálculos		X												X		X

**Anexo 3: Matriz de Trazabilidad Casos de Uso vs Sistema.**

Administración::CU-011: Administrar perfiles del sistema	MA::MA-001	MA::MA-002	MA::MA-003	MA::MA-004	MA::MA-005	MA::MA-006	MA::MA-007	MA::MA-008	MA::MA-009	MA::MA-010	MA::MA-011	MA::MA-012	MA::MA-013	MA::MA-014	MC::MC-001	MC::MC-002	MC::MC-003	MC::MC-004	MC::MC-005	MC::MC-006	MC::MC-007
Administración::CU-012: Asignar perfiles a usuarios del sistema																					
Administración::CU-025: Autorizar OR para SDL																					
Administración::CU-026: Definir plazos máximos																					
Carga de Información::CU-004: Cargar información mensual OR desde un archivo de datos																					
Carga de Información::CU-005: Cargar información diaria OR a través de un servicio B2B																					
Carga de Información::CU-006: Cargar información mensual OR a través de un servicio B2B																					
Carga de Información::CU-007: Validar información cargada por el OR																					
Carga de Información::CU-008: Generar indicador de calidad de la información del OR - Eliminado																					
Carga de Información::CU-010: Eliminar descargas de información del SUI																					
Carga de Información::CU-046: Cargar información mensual del SUI desde un archivo de datos																					
Carga de Información::CU-047: Cargar información del PAR desde un archivo de datos																					
Carga de Información::CU-048: Validar información del PAR																					
Carga de Información::CU-049: Validar la información mensual del SUI																					
Carga de Información::CU-058: Interfaz agentes comercial																					
Carga de Información::CU-059: Autenticar identidad usuario																					
Consultas y Reportes::CU-056: Generar reporte consolidado por elemento																					
Consultas y Reportes::CU-057: Generar reporte oficial de eventos																					
Procesar Información::CU-009: Verificar completitud de la información SUI																					
Procesar Información::CU-054: Calcular duración diaria de interrupciones																					
Procesar Información::CU-055: Calcular acumulado de interrupciones																					
Procesar Información::CU-060 Calcular EPD																					

## GUÍA PARA AUTORES

### Tipología de artículos para la revista

Cada uno de los artículos de la revista es evaluado por pares académicos con reconocida experiencia en el área a la cual pertenece el artículo enviado para su consideración en publicación. Los autores pueden presentar para publicación las siguientes tipologías, definidas por los organismos de indexación de las revistas científicas para reconocer la calidad de la publicación.

La revista otorga prelación para la publicación de artículos de tipo 1, 2 y 3.

1. Artículo de investigación científica y tecnológica. Documento que presenta, de manera detallada, los resultados originales de proyectos terminados de investigación. La estructura generalmente utilizada contiene cuatro apartes importantes: introducción, metodología, resultados y es.
  2. Artículo de reflexión. Documento que presenta resultados de investigación terminada desde una perspectiva analítica, interpretativa o crítica del autor, sobre un tema específico, recurriendo a fuentes originales.
  3. Artículo de revisión. Documento resultado de una investigación terminada donde se analizan, sistematizan e integran los resultados de investigaciones publicadas o no publicadas, sobre un campo en ciencia o tecnología, con el fin de dar cuenta de los avances y las tendencias de desarrollo. Se caracteriza por presentar una cuidadosa revisión bibliográfica de por lo menos 50 referencias.
- Otro tipo de artículos pueden ser considerados para publicación en la revista, dependiendo de su originalidad y aporte al conocimiento, según lo considere el comité científico, el comité editorial y el comité de árbitros de cada número. Entre ellos se cuentan:
4. Artículo corto. Documento breve que presenta resultados originales preliminares o parciales de una investigación científica o tecnológica, que por lo general requieren de una pronta difusión.
  5. Reporte de caso. Documento que presenta los resultados de un estudio sobre una situación particular con el fin de dar a conocer las experiencias técnicas y metodológicas consideradas en un caso específico. Incluye una revisión sistemática comentada de la literatura sobre casos análogos.

6. Revisión de tema. Documento resultado de la revisión crítica de la literatura sobre un tema en particular.
7. Cartas al editor. Posiciones críticas, analíticas o interpretativas sobre los documentos publicados en la revista que, a juicio del Comité Editorial, constituyen un aporte importante a la discusión del tema por parte de la comunidad científica de referencia.
8. Editorial. Documento escrito por el editor, un miembro del comité editorial o un investigador invitado, sobre orientaciones en el dominio temático de la revista.

### Envío de manuscritos

Los aspectos de forma deben seguir las normas básicas que la revista establece a la hora de redactar el informe o trabajo, en aspectos esenciales como los titulares, el tamaño de los párrafos, las citas bibliográficas, las nomenclaturas, la proporción de tablas y Figuras, la presentación de los resultados o las citas bibliográficas. En este caso también se recomienda a los autores que se ajusten a la normativa. Es necesario poner especial cuidado en la organización y la estructura del trabajo, el estilo de redacción, la presentación de los resultados en tablas y Figuras, y en la correspondencia entre las referencias citadas y las enumeradas al final del trabajo. Además, conviene no olvidar el ajuste del lenguaje a un entorno internacional, propio de la comunidad científica y profesional, más allá de los modismos propios del país o del área en la que se produzca el artículo.

Los autores deberán remitir los manuscritos de los artículos a través de la plataforma de la revista, a la cual se accede a través de <http://www.funlam.edu.co/lampsakos>, siguiendo las instrucciones que se indican en la plantilla oficial de la revista, consistente en los siguientes requerimientos mínimos:

#### Utilizar plantilla oficial de la revista.

Los manuscritos pueden ser escritos en español o inglés, cuartillas tamaño carta (21.59 cm × 27.94 cm) y márgenes de 2 cm en cada lado.

#### Máxima extensión del artículo.

7.000 palabras.

#### Formato de párrafos

Letra Arial, tamaño 10, espacio doble en columna sencilla.

#### Resumen.

Entre 150 y 250 palabras, en español e inglés. Incluir 3 a 5 palabras clave en orden alfabético, con su equivalencia en inglés.

#### Numeración de títulos y encabezados.

En listas multiniveles con números arábigos hasta el subnivel 3.

#### Cuerpo del manuscrito.

Incluir introducción, metodología y desarrollo evidenciando el aporte al área de ingeniería, es y trabajos futuros.

#### Citaciones y referencias.

Numeración bibliográfica según formato IEEE. Todas las citas en el manuscrito deben estar anotadas en las REFERENCIAS; en consecuencia, no deben existir REFERENCIAS aisladas que no estén citadas dentro del cuerpo del manuscrito.

### Proceso de evaluación y arbitraje

Los artículos para la revista *Lámpsakos* se someten a consideración del Comité de árbitros para cada edición. Al momento de enviar su manuscrito, la revista puede solicitarle sugerencias para la conformación del mencionado comité de evaluación, con la condición de que el proceso de dictamen para cada edición exige anonimato.

En primer lugar, los artículos recibidos serán objeto de una evaluación preliminar por parte de los miembros del Comité editorial, quienes determinarán la pertinencia de la publicación.

Una vez establecido que el artículo cumple con los requisitos temáticos, además de los requisitos formales indicados en estas instrucciones, será enviado a dos pares académicos externos, quienes determinarán en forma anónima: a) publicar sin cambios, b) publicar cuando se hayan cumplido correcciones menores, c) publicar una vez que se haya efectuado una revisión a fondo y d) rechazar. En caso de discrepancia entre ambos resultados, el texto será enviado a un tercer árbitro, cuya decisión definirá su publicación.

El proceso de dictamen se efectúa por pares académicos especialistas bajo la modalidad doble ciego. Cada par académico revisará la calidad, originalidad, relevancia y valor práctico del trabajo.

La notificación de aceptación o rechazo del artículo se realizará vía e-mail. El informe será emitido a los autores durante el mes siguiente al envío del manuscrito. Los resultados del proceso de dictamen académico serán inapelables en todos los casos.

En general, los aspectos de fondo más valorados serán la coherencia y la relevancia de la colaboración para el avance del conocimiento académico y profesional de la ciencia. Es necesario tener presente que cada tipología de trabajo tiene su propia estructura textual y conceptual, en función del aporte y de los posibles lectores.

### **Derechos de autor**

La revista *Lámpsakos* requiere a los autores que concedan la propiedad de sus derechos de autor, para que su artículo y materiales sean reproducidos, publicados, editados, fijados y comunicados y transmitidos públicamente en cualquier forma o medio, así como su distribución en el número de ejemplares que se requieran y su comunicación pública, en cada una de sus modalidades, incluida su puesta a disposición del público a través de medios electrónicos, ópticos o de cualquier otra tecnología para fines exclusivamente científicos, culturales, de difusión y sin fines de lucro. Cada artículo se acompaña de una *declaración de originalidad* en la que se especifique que no ha sido publicado y que no se someterá simultáneamente a otras publicaciones antes de conocer la decisión del comité editorial.

### **Originalidad y divulgación de artículos**

Los trabajos publicados en la revista, sin excepciones, se acogen a las normas del Copyleft y Creative Commons, ya que la publicación es de libre distribución para el conocimiento y podrá ser reproducida por cualquier medio, haciendo mención de la fuente. El autor o autores autorizan a la revista y a la Institución para editar y divulgar/publicar el artículo por cualquier medio nacional y/o internacional, impreso o electrónico

El autor o autores aceptan las políticas editoriales y los lineamientos de la guía. Los artículos son producto de un proceso de investigación de los autores y han sido valorados previamente por colegas expertos antes de ser presentados a publicación.

Las opiniones y afirmaciones que aparecen en la publicación reflejan exclusivamente los puntos de vista de sus autores y no comprometen necesariamente las políticas y pensamiento de la FUNLAM, la Facultad de Ingenierías o la revista *Lámpsakos*.

## GUIDE FOR AUTHORS

### Typology of articles for the journal

Each paper in the journal is evaluated by academic peers with recognized experience in the area on which the article belongs. Authors may submit their manuscripts for publication, according to the following typologies, defined by indexing agencies of scientific magazines that recognize the quality of the publications. (Publindex, Colciencias–Colombia)

The magazine gives priority to the publication of articles of type 1, 2 and 3.

1. Article of scientific and technological research. Document that presents, in detail, the original results of completed research projects. The structure generally contains four important parts: introduction, methodology, results and s.
2. Article of reflection. Document that presents finished research results from an analytical, interpretative or critical, on a specific topic, using original sources.
3. Review article. Document resulting from a completed investigation, which analyze, systematize and integrate the results of research published or unpublished, in a field of science or technology, in order of accounting for the progress and development trends. It is characterized by a careful literature review of at least 50 references.

Another kind of articles can be considered for publication in the magazine, depending on their originality and contribution knowledge according to the scientific committee and the editorial board of referees committee of each number. Among them are:

4. Short article. Brief document that presents original results preliminary or partial of a scientific or technological research, which usually requires a quick diffusion.
5. Case report. Document that presents the results of a study on a particular situation in order to make known the technical and methodological experiences considered in a specific case. Includes commented systematic review of the literature on analogous cases.
6. Topic Review. Document resulting from a critical review of the literature on a particular topic.
7. Letters to the Editor. Critical positions, analytical or interpretative on documents published in the magazine, which in the opinion of the Editorial Committee are an important contribution to the discussion of the topic by the scientific community of reference.
8. Editorial. Document written by the editor, editorial committee member or a guest researcher on guidelines in the thematic domain of the magazine.



## Requirements of manuscripts

The aspects of form must follow the basic rules that the journal has established for authors when writing their manuscripts, in essential aspects as the titles, the size of paragraphs, quotations, nomenclatures, the presentation of results, the proportion of tables and Figures. Authors should also conform to the rules of quotation of references. Special care is required in the organization and structure of the manuscript, as well as the writing style, the presentation of the results in tables and Figures, and the correspondence between the references cited and listed at the end of work. Moreover, authors must not forget about the language settings according to the international environment, typical of scientific and professional community.

Authors should refer their manuscripts of articles through the journal's platform, which may be accessed on the address <http://www.funlam.edu.co/lampsakos>, following the instructions in the official template magazine, consisting of the following minimum requirements:

### Use official magazine template

Manuscripts may be written in Spanish or English language, letter or A4 sized pages (21.59 cm x 27.94 cm) and 2 cm margins on each side.

### Maximum length of the article

7000 words.

### Paragraphs formatting

Arial letter, size 10, double-spaced in single column.

### Abstract

Between 150 and 250 words in both Spanish and English languages. Include 3-5 keywords in alphabetical order, with their equivalent in Spanish.

### Numbering of titles and headers

In multilevel lists with Arabic numerals to the sublevel 3.

### Body of manuscript

It includes Introduction, Development Methodology and evidencing the contribution to engineering, s and future work.

### Citations and bibliographic references

Numbering according to IEEE format. All citations in the manuscript must be listed in the references. There should not be isolated references that are cited in the body of the manuscript.

## Evaluation process and arbitration

The articles for the Journal "*Lámpsakos*" are subjected to consideration of the Committee of referees for each edition. When authors send their manuscript, the journal may solicit suggestions for the creation of that evaluation committee, based upon the fact that in the dictamination process for each edition requires anonymity.

First, the received articles will be subject to a preliminary assessment by the Editorial Committee members, who will determine the relevance of the publication.

Once established the Article complies with the thematic requirements in addition to the formal requirements indicated in these instructions, it will be sent out to external academic peers who determine on an anonymous basis: a) publish without changes, b) publish corrections they have complied with minor corrections, c) publish once having made a thorough review, d) reject. In case of discrepancy between the two results, the text will be sent to a third referee, whose decision will define its publication.

Dictamination process is carried out by academic peers specialists under double-blind mode. Each pair will review academic quality, originality, significance and practical value of the work.

Notification of acceptance or rejection of the article will be via e-mail. The report will be emitted into the authors during the month following the submission of the manuscript. The results of the academic opinion will be final in all cases.

In general, most valued aspects of background will be the consistency and relevance of collaboration in advancing academic and professional knowledge of science. It is necessary to remember that each type of job has its own textual and conceptual structure, depending on the amount and potential readers.

## Copyright

The journal *Lámpsakos* requires granting authors ownership of their copyrights, for their item and materials to be reproduced, published, edited, set and reported and publicly transmitted in any form means, and their distribution in the number of copies required, and public communication, in each of its forms

including the making available to the public through electronic, optical or other purpose any technology exclusively scientific, cultural, media and nonprofit. It has the letter of assignment of rights to do so.

### **Originality letter**

The works published in the magazine, without exception, are welcomed to the rules of Copyleft and Creative Commons, and that the publication is freely distributable for knowledge, and may be reproduced by any means making mention of the source. The author or authors authorize the magazine and the Institution to edit and disseminate/publish the article by any national and/or international, print or electronic.

Each article is accompanied by a statement of originality in specifying that has not been published and not simultaneously submitted to other publications before knowing the decision editorial committee.

Authors accept the editorial policies and guidelines of the guide. Their articles are the result of a research of the authors and has been previously assessed by expert colleagues before acceptance for publication.

Opinions and affirmations accepted for publication reflect exclusively the points of view of their authors and do not necessarily compromise the policies and thinking of either Fundación Universitaria Luis Amigó, the Faculty of Engineering or the journal *Lámpsakos*.