

# Inteligencia de negocios espacial para el análisis de los resultados de las pruebas para ingreso a la educación superior - SABER 11

Christiam Alejandro Niño-Peña<sup>#1</sup>, Gustavo Cáceres-Castellanos<sup>#2</sup>, Juan Sebastián González-Sanabria<sup>#3</sup>

<sup>#</sup> Escuela de Ingeniería de Sistemas y Computación, Universidad Pedagógica y Tecnología de Colombia, Av. Central del Norte – Edificio Central, Tunja-Boyacá, Colombia

<sup>1</sup> cristiam.nino@uptc.edu.co

<sup>2</sup> gustavo.caceres@uptc.edu.co

<sup>3</sup> juanesebastian.gonzalez@uptc.edu.co

**Abstract**— In this work, there is an analysis obtained from applying spatial business intelligence on the results of Pruebas Saber 11 tests applied by the ICFES (Acronym of Colombian Institute for the Promotion of Higher Education). The people evaluated, the educational institutes, the socio-economic conditions and the municipalities were taken into account for the analysis. For a better analysis, some applicative were used to manipulate data to obtain a spatial data warehouse and apply spatial data mining techniques, in order to make a spatial analysis that can be visualized on maps.

**Keywords** – spatial analysis, spatial business intelligence, spatial data mining, spatial datawarehouse.

**Resumen**— En el trabajo se presenta el análisis obtenido de aplicar inteligencia de negocios espacial sobre los resultados obtenidos en la Pruebas Saber 11 aplicadas por el Instituto Colombiano para el Fomento de la Educación Superior (ICFES). Para el análisis, se tuvieron en cuenta los evaluados, las instituciones educativas, las condiciones socioeconómicas y los municipios. Con el fin de obtener un mejor análisis, se usaron aplicativos que permitieron manipular los datos para obtener la bodega de datos espacial y aplicar técnicas de minería de datos espacial con el fin de realizar un análisis espacial que permitiera ser visualizado en mapas.

**Palabras clave** – análisis espacial, bodega de datos espacial, inteligencia de negocios espacial, minería de datos espacial.

## I. INTRODUCCIÓN

La información almacenada en una organización a través del tiempo puede contener datos, de cualquier tipo, relacionados con procesos en diversas temáticas que requieren ser analizados históricamente para poder comprobar el cumplimiento de las metas propuestas. Para obtener este análisis, se pueden aplicar técnicas de inteligencia de negocios, las cuales resultan bastante útiles y proporcionan un marco de referencia para la toma de decisiones. Sin embargo, en ocasiones, estos análisis son insuficientes cuando se requiere obtener información que comprometa un lugar o marco geográfico [1].

Como alternativa a dicha problemática surge la inteligencia de negocios espacial, la cual realiza una serie de etapas que comienzan con la recolección de los datos transaccionales históricos con su componente espacial, luego se procede a la extracción, transformación y carga de los datos, posteriormente se analizan y diseñan las dimensiones y medidas para el modelado del cubo espacial,

aplicando minería de datos espacial con el fin de brindar los elementos suficientes para la toma de decisiones [2].

El Instituto Colombiano para el Fomento de la Educación Superior (ICFES) aplica anualmente las pruebas Saber 11, con los objetivos de “a) Seleccionar estudiantes para la educación superior, b). Monitorear la calidad de la formación que ofrecen los establecimientos de educación media y c). Producir información para la estimación del valor agregado de la educación superior” [3]. Por lo anterior, y para obtener mejores resultados en el análisis, se requiere hacer énfasis en el tratamiento de los datos utilizando el componente espacial.

En el presente trabajo se presenta la implementación de inteligencia de negocios espacial para el análisis de resultados de las pruebas Saber 11 del ICFES. Se inició el proceso con de seleccionar una herramienta software completa que permita aplicar inteligencia de negocios espacial. Posteriormente, se diseñó y creo la bodega de datos y cubo espacial, lo cual brinda el insumo de información para el tratamiento de la información. Paso seguido se seleccionó y aplicó el algoritmo de minería de datos espacial para obtener el conocimiento asociado. Se finaliza con la implementación de un visor de información espacial que permite manipular y observar los resultados de las medidas y dimensiones para la toma de decisiones espacial de forma clara y completa.

En el proyecto se convierte en un elemento que permite tomar decisiones sobre diversos factores mediante una representación espacial en mapas para facilitar el análisis del comportamiento de los resultados de la aplicación de las pruebas, donde antes solo se tenía una visión cuantitativa en tablas y gráficos estadísticos.

## II. METODOLOGÍA

Considerando que el trabajo se enmarca en Inteligencia de Negocios Geoespacial, se tomó como base para su desarrollo la metodología planteada por Kimball, ajustada a datos espaciales, como se presenta a continuación:

- **Planificación:** En esta fase se define el propósito principal del proyecto, se determinan cuáles son sus objetivos específicos y su alcance, adicionalmente se da una aproximación inicial a las necesidades que se intentan suplir con su desarrollo.

- **Análisis de Requisitos:** En esta fase se debe indagar sobre el negocio y los procesos de estudio, recopilar y analizar los reportes e informes que arroja cada proceso, entrevistar a los empleados y conocer la terminología del

negocio. A partir de los requisitos analizados se puede construir la matriz de procesos/dimensiones, esta matriz facilita la priorización de los requisitos o procesos más críticos.

- **Modelado Dimensional:** A partir de la matriz de procesos/dimensiones, se construye el modelo dimensional. El primer paso para realizar la construcción del modelo dimensional, es elegir el proceso de negocio, luego se establece el nivel de granularidad, es decir el nivel de detalle dependiendo de los requisitos del negocio. Posteriormente, se eligen las dimensiones apoyándose en el nivel de granularidad elegido y la matriz de procesos/dimensiones. El siguiente paso es identificar las tablas de hechos y medidas. Teniendo lo anterior, se construye un gráfico denominado modelo multidimensional de alto nivel. Luego se identifican los atributos de las tablas de dimensiones y las tablas de hecho para construir el modelo dimensional detallado.

- **Diseño físico:** En esta fase se intentan contestar preguntas como las siguientes: ¿Qué hardware y software se necesita para el desarrollo?, ¿Qué necesitan instalar los equipos cliente para acceder al sistema?, ¿Cómo convertir el modelo de datos lógico en un modelo de datos físico en una base de datos?, ¿Cómo se debe configurar el sistema?

- **Diseño del sistema de Extracción, transformación y carga (ETL):** En esta fase se debe diseñar y definir como se realizará el proceso de extracción, transformación y carga de datos. Para almacenar los datos de las fuentes originales en la bodega de datos.

- **Especificación y desarrollo de aplicaciones de Inteligencia de Negocios:** En esta fase se busca brindar a los usuarios una forma más fácil, rápida y estructurada para acceder a los datos almacenados por medio de aplicaciones de inteligencia de negocios. Con dichas aplicaciones se busca proporcionar informes, reportes y herramientas de análisis de datos. Dichos informes pueden ser simples y estáticos o implementar modelos matemáticos de minería de datos.

### III. RESULTADOS

#### A. Planeación del proyecto

Para el desarrollo del trabajo se escogieron las pruebas Saber 11, donde además se contó con el aval del funcionario de la Oficina Asesora de Gestión de Proyectos de Investigación el cual apoyó el proceso de realización del trabajo de investigación.

#### B. Análisis de requisitos

Se analizaron las necesidades del ICFES en cuanto a manejo de la información. Es importante destacar que este Instituto sugiere que los investigadores interesados estudien las variables y recomienden algunos índices, indicadores o relacionados dependiendo del tipo de estudio a realizar, además hoy en día no existe ningún estudio, aplicación y/o herramienta que brinde apoyo a la toma de decisiones de manera espacial en la Institución [4].

El desarrollo inicial del proyecto fue planteado buscando dar respuesta a las necesidades del usuario. Para ello se determinó un conjunto de preguntas, con el propósito de que el sistema ofrezca las respectivas respuestas, las cuales se enuncian a continuación:

- ¿Cuál es la cantidad de evaluados en 2013 por calendario A, mixto, oficial, jornada mañana de las instituciones educativas en áreas de núcleo común por departamento y municipio según análisis espacial?
- ¿Cuál fue el mejor departamento de Colombia en obtener los puntajes más altos en los resultados de las pruebas saber 11 y al mismo tiempo cual fue el más bajo en el año 2013 representados espacialmente?
- ¿Cuáles fueron los máximos puntajes por área en los municipios de un departamento específico del año 2013 según análisis espacial?

#### C. Modelamiento dimensional

En el desarrollo de la bodega de datos espacial se utilizó un modelo en copo de nieve, detallando de forma espacial los municipios y departamentos de las instituciones educativas.

Se definieron las dimensiones y medidas del datamart como se ilustra en la Fig. 1.

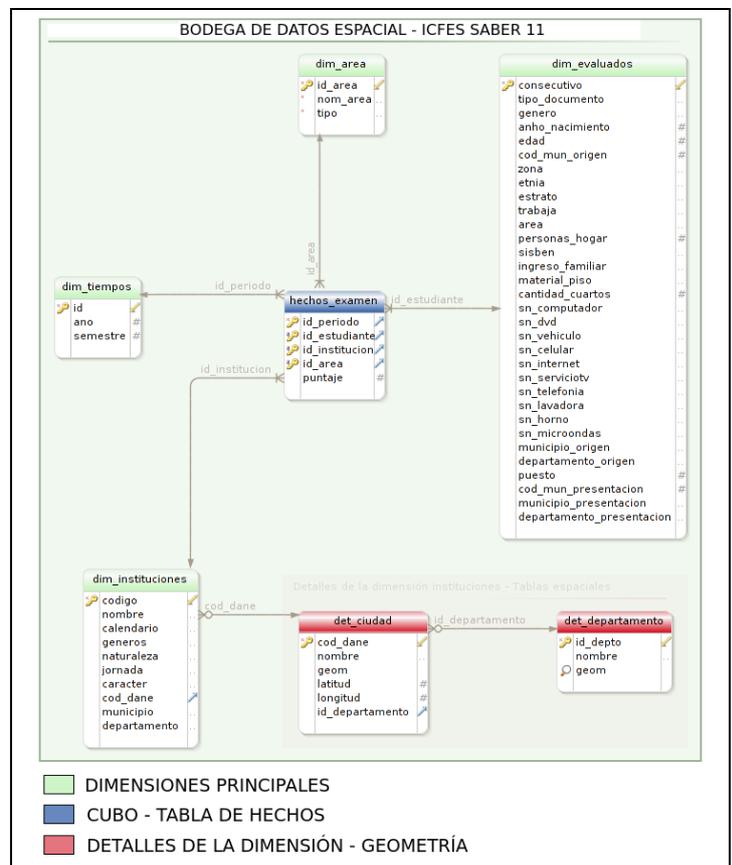


Fig. 1. Datamart espacial ICFES Saber 11.

#### D. Selección de herramientas

Para aplicar inteligencia de negocios espacial, se deben utilizar herramientas que permitan el manejo adecuado de los datos de manera que sea posible obtener resultados que favorezcan el cumplimiento de los objetivos y metas en una organización [1].

Las herramientas para los Sistemas de Información Geográfica (SIG) y para el procesamiento analítico en línea espacial (Spatial OnLine Analytical Processing-SOLAP), desempeñan un papel fundamental en el análisis de la información mediante gráficos, mapas, tablas e informes [2]. El estudio de caracterización presentado a continuación

permitió comparar las herramientas para brindar elementos de juicio que coadyuvaran a decidir cuál es la más adecuada para la implementación de Inteligencia de Negocios Espacial, la Tabla I expone la caracterización de los SIG y en la Tabla II se presenta lo propio para las herramientas SOLAP.

TABLA I  
CARACTERIZACIÓN DE HERRAMIENTAS SIG

Sistema	Licencia	Versión	Plataforma
ArcGIS	Privativa	10.1	Escritorio, Web, Móvil
GRASS	GPL V2	6.4, 7.0	Escritorio
ILWIS	GPL	3.8	Escritorio
TransCAD	Privativa	2014	Escritorio, Web
ERDAS	Privativa	13.0.2	Escritorio
I3Geo	GPL	4.7	Web
QGIS	GPL	2.0.1	Escritorio, Web, Móvil
GvSIG	GPL	2.1	Escritorio, Móvil
JMap	Privativa	5.0	Escritorio, Web, Móvil
uDig	EPL, BSD	1.2.1	Escritorio
Saga	GPL, LGPL	2.0	Escritorio
OpenEV	LGPL	1.8.0	Escritorio

TABLA II  
CARACTERIZACIÓN DE HERRAMIENTAS SOLAP

Sistema	Licencia	Versión	Plataforma
GeoKettle	LGPL	2.5	Escritorio
GeoMondrian	EPL	1.0	Web
Schema Workbench	GPL	1.0	Escritorio
Knime	GPL	1.0	Escritorio
Weka	GPL	3.6.6	Escritorio
Saiku	GPL	1.0	Web
Solap Layers	GPL	2.0	Web

Posterior a la anterior caracterización y teniendo en cuenta los demás aspectos implícitos en el desarrollo, se eligieron las siguientes herramientas libres:

- Sistema Gestor de Base de Datos Espacial: Postgres+Postgis [5, 6].
- Software para construir el Cubo de Datos: GeoKettle [7].
- Servidor SOLAP Geospacial: GeoMondrian [8].
- Software Minería de datos Espacial: GeoKnime [9].
- Software de visualización Geospacial BI: SOLAP Layers [10].

Cabe resaltar que actualmente no se existe una plataforma o un sistema que integre cada las herramientas mencionadas para aplicar inteligencia de negocios espacial [11].

### E. Diseño arquitectura

La arquitectura diseñada para el desarrollo de la bodega de datos espacial de las pruebas ICFES Saber 11 se muestra en la Fig. 2.

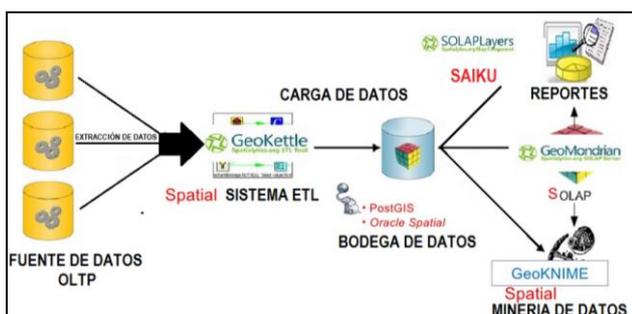


Fig. 2. Arquitectura técnica bodega de datos espacial. Basado en [12].

### F. Especificación y desarrollo de aplicaciones de Inteligencia de Negocios

Teniendo en cuenta los requisitos de información y preguntas a responder con el sistema se determina que la tarea de minería a realizar es descriptiva y específicamente de agrupamiento.

- Selección del algoritmo. El algoritmo más adecuado en la aplicación del proyecto es K-MEANS espacial, por su eficiencia, integración de elementos alfanuméricos y espaciales, su capacidad para encontrar valores permitiendo representarlos espacialmente, la posibilidad de utilizar los resultados de los puntajes de las pruebas Saber 11 y que permite analizar patrones espaciales de los casos según la media de la distancia entre puntos; lo anterior a través del agrupamiento de puntajes para establecer clústeres de evaluación según la clasificación del ICFES que son: alto (71-100), medio (31-70) y bajo (0-30) para el año 2013.
- Aplicación de minería de datos espacial. Se realizaron una serie de pasos que se detallan en la Fig. 3 e incluyen la aplicación del algoritmo K-Means espacial.

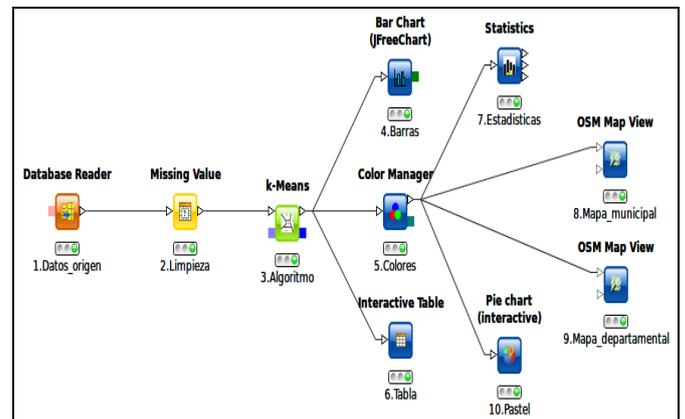


Fig. 3. Aplicación de minería de datos espacial.

- Resultados encontrados. Los resultados de las diferentes áreas de evaluación se representan en diagramas de pastel y mapas, donde se clasificaron: clúster 2: color verde; clúster 1: color amarillo y clúster 0: color rojo.

El conocimiento encontrado se estructuró en los 3 clústeres que serán representados por la latitud y longitud en el mapa por resultados, estos resultados se clasificaron según la tabla de puntajes en alto, medio y bajo explicados anteriormente, a continuación, se muestra el resultado hallado para el área de biología en la Fig. 4.

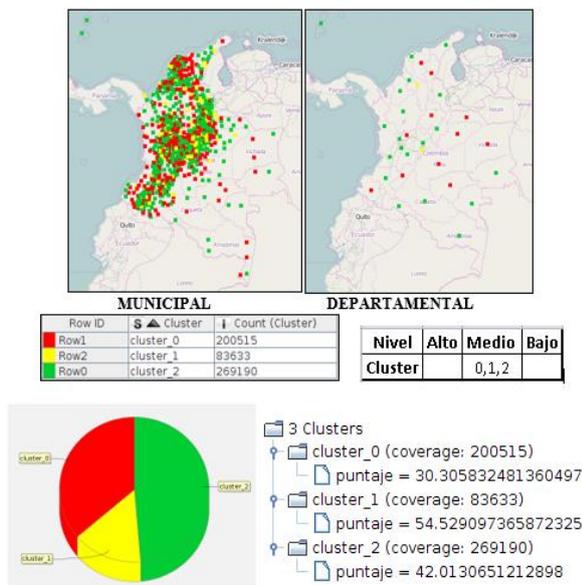


Fig. 4. Resultados de minería de datos para el área de Biología año 2013.

G. Resultados a las preguntas objetivo

- 1) ¿Cuál es la cantidad de evaluados en 2013 por calendario A, mixto, oficial, jornada mañana de las instituciones educativas en áreas de núcleo común por departamento y municipio según análisis espacial?  
 Respuesta: El resultado encontrado en la Fig. 5 muestra que la ciudad en tener más evaluados según las características de la pregunta es Bogotá, y los departamentos Antioquia y Valle del Cauca.

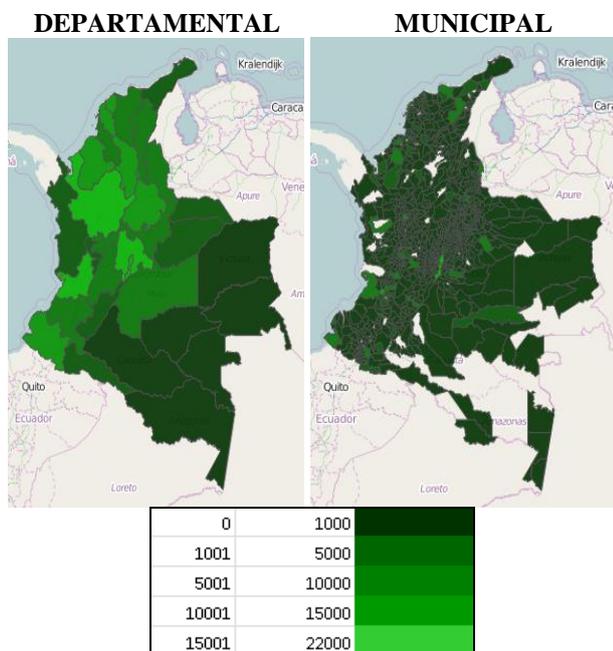


Fig. 5. Cantidad de evaluados en 2013 por: calendario A, mixto, oficial, jornada mañana de las instituciones educativas en áreas de núcleo común por departamento y municipio según análisis espacial.

- 2) ¿Cuál fue el mejor departamento de Colombia en obtener los puntajes más altos en los resultados de las pruebas saber 11 y al mismo tiempo cual fue el más bajo en el año 2013 representados espacialmente?

Respuesta: El mejor departamento fue Antioquia y el peor fue Vaupés como se muestra en la Fig. 6.

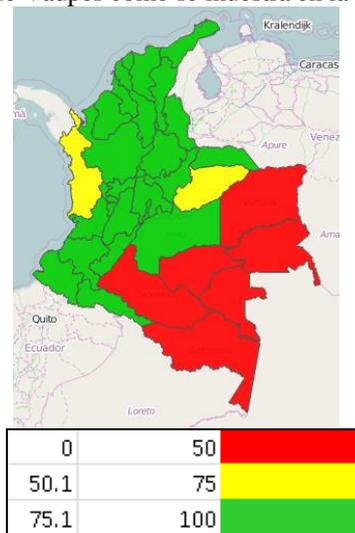


Fig. 6. Mejor departamento de Colombia en obtener los puntajes más altos en los resultados de las pruebas saber 11 y al mismo tiempo cual fue el más bajo en el año 2013 representados espacialmente.

- 3) ¿Cuáles fueron los máximos puntajes del área de matemáticas en los municipios del departamento de Boyacá del año 2013 según análisis espacial?  
 Respuesta: En los municipios donde se obtuvo el mayor puntaje de matemáticas fue en Paipa, Duitama, Sogamoso, Puerto Boyacá y Tunja como se muestra en la Fig. 7.

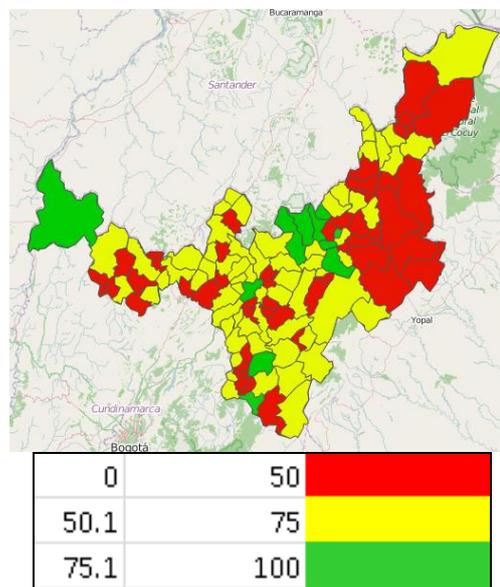


Fig. 7. Máximos puntajes del área de matemáticas en los municipios del departamento de Boyacá del año 2013 según análisis espacial.

IV. CONCLUSIONES

A hoy, se han desarrollado una cantidad limitada de proyectos relacionados con la inteligencia de negocios espaciales, esto por ser una tecnología un tanto nueva y de compleja exploración, debido a que involucra nuevos tipos de datos, técnicas de aplicación y un alto manejo de procesamiento y memoria. Por ésta razón, el almacenamiento espacial ha evolucionado incorporando nuevas formas de obtener el tipo de dato espacial y mejorando la distribución de las tablas e índices para la

recuperación de los mismos, a su vez los modelos y consultas espaciales incrementan el rendimiento e integridad en la recuperación de la información, teniendo en cuenta que el uso de estas consultas exige más y mejores procedimientos. En concordancia, el surgimiento de nuevos algoritmos para extraer la información espacial hace que la minería de datos espacial desarrolle técnicas de transformación de los datos para que cumplan con los requerimientos exigidos.

El almacenamiento espacial evidencia la construcción de dimensiones y medidas según esquemas y reglas personalizadas, construcción de modelos y manejo de niveles de granularidad, además de la representación de objetos con el uso de índices especializados en el manejo de las transacciones.

En el desarrollo de la bodega espacial el dato de tipo geométrico fue importante para relacionar los datos alfanuméricos con un lugar específico y además poder ubicarlo en un mapa o espacio cartográfico, razón por la cual se tuvo que tener en cuenta su proyección y definición espacial.

Actualmente existe una gran variedad de técnicas y algoritmos orientados a la minería de datos, estos algoritmos poco a poco se han venido adaptando al uso espacial de forma que permitan la aplicación de minería de datos espaciales y descubrir conocimiento que antes era imposible o difícil de encontrar, esta razón a hecho que nazcan proyectos como GeoKnime que, aunque están en proceso de maduración son útiles a la hora de utilizar este tipo de información espacial.

La aplicación de inteligencia de negocios espacial en las pruebas SABER 11 del ICFES coadyuva a enfocar los esfuerzos en donde realmente se requieren, esto al poder mostrar el conocimiento sobre las áreas geográficas, y de

esta forma aplicar proyectos que beneficien al país en su desarrollo académico.

#### TRABAJOS FUTUROS

Se propone la implementación de diversas métricas, así como el uso de otras plataformas libre para inteligencia de negocios espacial que se vayan desarrollando, lo anterior teniendo en cuenta que, por el auge actual de la temática, continuamente se presentan nuevas tecnologías que apuntan a facilitar los procesos que ésta área conlleva.

#### REFERENCIAS

- [1] Sugumaran, R. and J. Degroote, "Spatial Decision Support Systems - Principles and Practices", in CRC Press, pp. 507, 2011
- [2] Flacke, J., "Spatial Decision Support Systems – Introduction to Concepts and Requirements". Web-GIS and Spatial Data Infrastructure, 2012.
- [3] Gobierno de Colombia. Decreto 869 de 2010.
- [4] ICFES. "Repositorio FTP ICFES" 2014. [online]. Available from: <ftp://ftp.icfes.gov.co>.
- [5] Postgresql. "Postgresql" 2013. [online]. Available from: <http://www.postgresql.org.es/>.
- [6] Postgis. "Postgis" 2013. [online]. Available from: <http://postgis.net/>.
- [7] Spatialytics.org. "Geokettle" 2013. [online]. Available from: <http://www.spatialytics.org/projects/geokettle/>.
- [8] Spatialytics.org. "Geomondrian" 2013. [online]. Available from: <http://www.spatialytics.org/projects/geomondrian/>.
- [9] AG. "Geoknime" 2013. [online]. Available from: <http://www.knime.org>.
- [10] tbadard. "Solaplayers" 2010. [online]. Available from: <http://sourceforge.net/projects/spatialytics/>.
- [11] Badard, T., "Open source Geospatial Business Intelligence in action with GeoMondrian and SOLAPLayers!" 2010. [online]. Available from: <http://www.slideshare.net/tbadard/open-source-geospatial-business-intelligence-geobi-definition-architectures-projects-challenges-and-outlooks-8810753>.
- [12] Lamas, A., et al., "Creación de un módulo espacial OLAP para Saiku", in VIII Jornadas de SIG Libre de Girona, pp. 9, 2013.