

## Cómo citar el artículo

Blanco Rojas, T., Archila Córdoba, D. M. & Ballesteros Ricaurte, J. A. (2016). Gestión de datos obtenidos desde redes sociales aplicando *Business Intelligence Engineering Process*. *Revista Virtual Universidad Católica del Norte*, 49, 72-91. Recuperado de <http://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/798/1318>

# Gestión de datos obtenidos desde redes sociales aplicando *Business Intelligence Engineering Process*\*

## **Tatiana Blanco Rojas**

Ingeniera de Sistemas y Computación  
Universidad Pedagógica y Tecnológica de Colombia  
[tatianablancorojas@gmail.com](mailto:tatianablancorojas@gmail.com)

## **Diana Milena Archila Córdoba**

Ingeniera de Sistemas y Computación  
Universidad Pedagógica y Tecnológica de Colombia  
[diana.archila.cordoba@gmail.com](mailto:diana.archila.cordoba@gmail.com)

## **Javier Antonio Ballesteros-Ricaurte**

Ingeniero de Sistemas  
Magíster en Ciencias Computacionales  
Docente asociado de la Escuela de Ingeniería de Sistemas y Computación  
Universidad Pedagógica y Tecnológica de Colombia  
[javier.ballesteros@uptc.edu.co](mailto:javier.ballesteros@uptc.edu.co)

**Recibido:** 30 de junio de 2015.

**Evaluado:** 2 de agosto de 2016.

**Aprobado:** 8 de agosto de 2016.

**Tipo de artículo:** investigación científica y tecnológica.

\* Artículo de investigación resultado del proyecto *Sistema de Información para obtener datos mediante redes sociales usando Business Intelligence Engineering Process*, asociado a la participación en grupo de investigación - GIMI. Fecha de inicio: septiembre de 2013. Fecha de finalización: septiembre de 2014.

## Resumen

Desde hace algún tiempo, las redes sociales han tomado mucha fuerza dentro de las organizaciones e instituciones, llegando al punto de investigar los datos contenidos en estos medios de comunicación para la solución de diversas necesidades. En la actualidad, la Universidad Pedagógica y Tecnológica de Colombia afronta una situación problemática basada en la falta de comunicación con los egresados de la misma; por tal razón, en el presente artículo de investigación se presentan los resultados obtenidos en el proceso de extracción de datos desde redes sociales aplicando la metodología *Business Intelligence Engineering Process* (BIEP).

## Palabras clave

*Business Intelligence Engineering Process*, Egresados, Facebook, Inteligencia de negocios, Minería de datos, Redes sociales.

### Management of the Data Obtained from Social Networks by Applying Business Intelligence Engineering Process (BIEP)

## Abstract

In recent times social networks has become very important within organizations and institutions to the point of researching the data contained in this communication media for solving different needs. Currently the Pedagogical and Technological University of Colombia has

a problematic situation derived from the lack of communication with its graduates; therefore, in this research article we present the results obtained from the process of collecting data from social networks by using the BIEP methodology.

## Keywords

BIEP, Graduates, Facebook, Business intelligence, Data mining, Social networks.

### Gestion de données obtenues de réseaux sociaux en appliquant *Business Intelligence Engineering Process* (BIEP)

## Résumé

Depuis un temps, les réseaux sociaux ont devenu très importants chez les organisations et les institutions, à tel point qu'on recherche les données dérivées de ces moyens de communication pour solutionner des différentes nécessités. Dans l'actualité, l'Université Pédagogique et Technologique de la Colombie a une situation problématique qui se base sur le manque de communication avec ses diplômés ; dans cet article on présente les résultats obtenus dans le processus d'extraction de données des réseaux sociaux en appliquant la méthodologie BIEP.

## Mots-clés

BIEP, Diplômés, Facebook, Informatique décisionnelle, Extraction de données, Réseaux sociaux.

## Introducción

La nueva era de tecnología ha obligado a las personas, instituciones y organizaciones a estar informadas; por tal razón, es frecuente ver cómo la gente cuenta con cualquier clase de dispositivo que pueda tener acceso a internet para estar en contacto y actualizado. Este contexto ha llevado a que los medios de comunicación masiva sean populares y tengan mucha demanda por parte de los usuarios para obtener datos que beneficien, de una u otra forma, a grupos o individuos que buscan información en común. Las fuentes más usadas en estos días para actualizarse e informarse son las llamadas redes sociales, pues son un medio de comunicación innovador con grandes contenidos de información.

La Universidad Pedagógica y Tecnológica de Colombia (UPTC) es una institución educativa que cuenta con una población de egresados; sin embargo, el sistema que maneja la información de dicha población no contiene los datos actualizados a la fecha. Por otro lado, ese fenómeno de desactualización también afecta a la Escuela de Ingeniería de Sistemas y Computación de la UPTC – seccional Tunja: allí se percibe ausencia de una base de datos estable y actualizada, pues la información que se maneja de la población en estudio en algunos casos no es real y tampoco es administrada de manera correcta para realizar contactos continuos y permanentes,

relacionados con aspectos sociales, académicos y culturales desarrollados por la Escuela y la Universidad.

Durante el desarrollo del presente artículo, en la segunda sección (Marco teórico) están contenidos los conceptos básicos que hacen parte de la investigación; en la tercera se describe la metodología de desarrollo junto a todos sus pasos; la cuarta describe los resultados, mostrándolos en dos partes importantes; la quinta muestra los reportes generados por consultas y los reportes personalizados; y en la sexta se definen las conclusiones de la investigación.

## **Marco teórico**

En esta sección se presentan las teorías más importantes tenidas en cuenta para el desarrollo de la investigación.

### **Las redes sociales**

Con el rápido crecimiento de las cuentas de usuario en las redes sociales se puede observar que estas son un foco inmenso de todo tipo de información, entre la que se encuentran gustos, sentimientos, emociones, eventos, lugares y demás (Grau, 2009). Con estos datos las organizaciones pueden estudiar las necesidades de los usuarios a través de metodologías y herramientas especializadas, encontrando allí soluciones para publicidad, mercadeo, comunicación, presupuesto, manejo de bases de datos y *marketing* (I-Hsien, Hui-Ju & Pei-Shan, 2009).

Las redes sociales contienen datos indispensables para las organizaciones, pero en algunas ocasiones no son tenidas en cuenta; de aquí se deriva el problema de que no se usen metodologías y arquitecturas adecuadas para transformar datos en información útil (Caldevilla Domínguez, 2010).

### **NoSQL**

Las bases de datos NoSQL (*Not Only SQL* – No solo SQL) surgen como una solución a los constantes requerimientos de procesamiento y análisis a gran escala de enormes cantidades de datos, para los cuales los sistemas tradicionales de base de datos son insuficientes. NoSQL se enfoca en sistemas de almacenamiento de información que no cumplen con el esquema entidad-relación (Martín et al., 2013). La gran flexibilidad que ofrece, así como las posibilidades que brinda desde el punto de vista de la optimización en sus diseños, la convierten en una atractiva variante a tener en cuenta para la administración de información (Gracia del Busto & Yanes Enríquez, 2012).

## BIEP: Metodología para el desarrollo de sistemas orientados a inteligencia de negocios

La presente investigación se enfoca en la aplicación de una metodología de ingeniería de software para el desarrollo de un sistema de inteligencia de negocios (*Business Intelligence-BI*), la cual es conocida como "Proceso de ingeniería en inteligencia de negocios" (en inglés: *Business Intelligence Engineering Process - BIEP*) y está compuesta por dos elementos: las fases y los flujos de trabajo, transversales a los elementos que componen un sistema de inteligencia de negocios. Esta metodología se fundamenta en el proceso unificado. En la parte de integración de datos, BIEP se basa en la metodología de proceso de ingeniería para el desarrollo de bodegas de datos o *Data Warehouse Engineering Process (DWEP)*. En el análisis de datos se utilizan las técnicas OLAP (Jadav & Panchal, 2012) o de minería de datos (*Data Mining – DM*) (Herrera Osorio, 2011).

En ese orden de ideas, BIEP contiene ocho flujos de trabajo, cinco de ellos derivados del proceso unificado: requerimiento, análisis, diseño, implementación y pruebas. Dos de ellos provienen del DWEP: el mantenimiento y las revisiones posdesarrollo. Esta Metodología también contiene las mismas fases desarrolladas en el proceso unificado y en

DWEP, como las mostradas en la Tabla 1.

**Tabla 1.** Fases de la metodología BIEP

Fase	Descripción
Inicio	Define el límite y alcance del proyecto. Objetivo principal: proveer una visión general del proyecto de BI que se pretende desarrollar.
Elaboración	Plantea el plan de proyecto, características que le identifican y línea base de la arquitectura. Con la agrupación de estos elementos se obtiene la arquitectura base del proyecto.
Construcción	Se lleva a cabo el desarrollo del proyecto con el fin de entregar un prototipo inicial al usuario o cliente.
Transición	Implantación del proyecto en el entorno en el que operará haciendo partícipes a los usuarios. Su meta es entregarles la versión final.

## Metodología de desarrollo

Tomando como base la metodología BIEP se desarrollan cuatro (4) fases compuestas de actividades, con el fin de cumplir los objetivos propuestos.

## Fase de inicio

Con el fin de llevar a cabo el proceso para determinar la influencia de las redes sociales en la obtención de información se definen las siguientes actividades:

### a. Flujo de requerimientos

El resultado final de este flujo de trabajo es conocer los requerimientos del proceso de inteligencia de negocios, utilizando el modelo de caso de uso del cual se desprenden las siguientes labores:

- Establecer los objetivos del negocio.
- Diseñar un *Data Mart* constituido por los datos de los egresados de la Escuela de Ingeniería de Sistemas y Computación de la UPTC, obtenidos de las redes sociales Facebook, Twitter y LinkedIn.
- Elaborar reportes a través de consultas diseñadas con los elementos internos de la base de datos.
- Objetivo de minería de datos: realizar un análisis y clasificación de los datos obtenidos de egresados en la minería de datos en las redes sociales *Facebook, Twitter y LinkedIn*.

### b. Flujo de análisis

Este flujo de trabajo tiene como objetivo el refinamiento y estructuración de los requerimientos obtenidos en el flujo de requerimientos, para lo cual se desarrollan las siguientes actividades:

**Recopilación de datos:** La identificación de las fuentes de datos requiere saber qué datos están disponibles para alimentar la bodega de datos. Para este proceso, después de conocer el problema, se contactó a algunos egresados a través de la red social Facebook, quienes de manera voluntaria respondieron una serie de preguntas de tipo personal, académico y laboral. También se procedió a realizar búsquedas manuales en LinkedIn: se encontró información más completa y veraz en esta red por cuanto presenta información de carácter profesional. Y sumado a lo anterior, se realizó la búsqueda en Twitter para confrontar los datos obtenidos con las otras dos redes.

Una vez recopilada esta información se procedió a compararla con la base de datos actual de la población en estudio, lo que arrojó como resultado el hecho de que la Escuela de Ingeniería de Sistemas y Computación tenía información desactualizada de sus egresados. Por todo lo anterior, las principales fuentes de información en las cuales se basa esta investigación son las siguientes:

- Redes sociales: Facebook, Twitter y LinkedIn
- Base de datos de egresados existente
- Encuesta a egresados

**Revisión de fuentes de datos:** consiste en el diseño de un modelo que representa una estructura más organizada de las fuentes de datos, basado en el modelo relacional; en él se pueden observar las clases y sus respectivas asociaciones.

**Modelo entidad-relación:** en la figura 1 se tienen cinco entidades, cada una de las cuales representa una fuente de información (“Egresados”, “AreasInteres”, “CursoAptitudesConocimientos”, “TitulosAcademicos” y “ExperienciaLaboral”); también se representan las relaciones de modularidad que existen entre ellas.

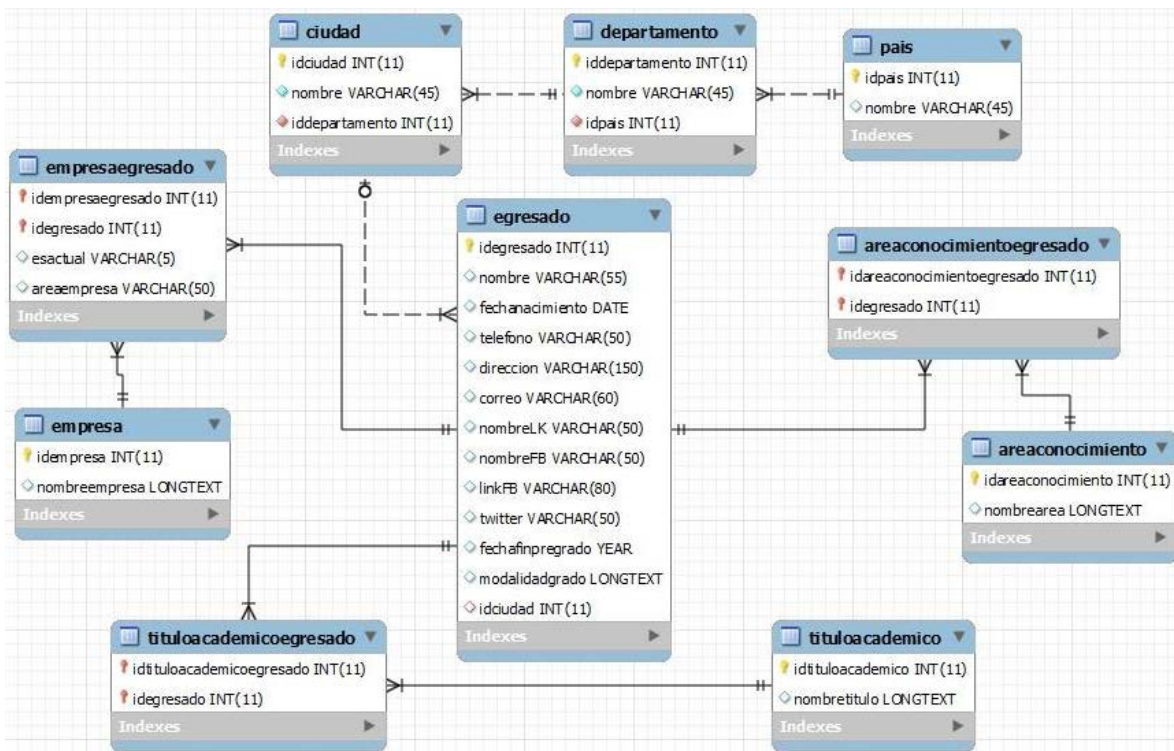


Figura 1. Modelo entidad – relación

**Diagrama de caso de uso:** en la figura 2 se muestra este diagrama para el desarrollo de la investigación, en el que se determinan los requisitos funcionales del sistema desarrollado.

- Actores
  - o Investigadores
  - o Escuela de Ingeniería de Sistemas y Computación
  - o Egresado

## - Caso de uso

- Buscar y organizar información: desarrollado por los investigadores, consiste en la realización de procesos de extracción, transformación y carga de datos.
- Hacer consultas: a partir de los datos existentes en la base de datos, se diseñan consultas para comprobar el funcionamiento de la base de datos y su consistencia.
- Generar reportes: con base en las consultas diseñadas, se usa la herramienta para generar reportes que permiten caracterizar el comportamiento de los datos.
- Plantear proyectos y estrategias de mejora: se establecen trabajos futuros que ayuden a optimizar el proceso de migración de datos.



Figura 2. Diagrama de caso de uso

**Descripción de los datos:** dentro de la recopilación de datos que se efectuó en las redes sociales en estudio (Facebook, Twitter y LinkedIn), se tuvieron en cuenta los siguientes parámetros a extraer con el fin de elaborar un historial de la muestra de egresados. Cabe aclarar que no todos los fueron encontrados, debido a la falta de actualización por parte de la población implicada en los perfiles de las redes sociales anteriormente nombradas. Los parámetros extraídos fueron los siguientes:

- Especialidades

- Experiencia
- Educación
- Datos de nacimiento
- Teléfono
- Ciudad
- Dirección
- Nombre del egresado
- Aptitudes y conocimientos
- Reconocimientos y premios
  
- Proyectos
  
- Certificaciones y cursos
  
- Idiomas
  
- Publicaciones
  
- Correo electrónico
  
- Nombre de usuario en LinkedIn
- Nombre de usuario en Facebook
- Vínculo en Facebook
  
- Twitter
  
- Fecha de graduación del pregrado
- Modalidad de grado
- Empresa actual
- Intereses

Para documentar el sistema de operaciones existente que alimenta la bodega de datos se hizo uso de los siguientes diagramas:

**Esquema conceptual de origen de datos (*Source Conceptual Schema, SCS*):** este esquema se deriva del modelo entidad-relación y detalla a fondo las relaciones existentes entre cada clase (figura 3). En la investigación se tiene que las redes sociales poseen dentro de sí varios usuarios que son egresados de la Escuela de Ingeniería de Sistemas y Computación. Las redes sociales deben conectarse a bases de datos híbridas (SQL-NoSQL) para cargar los perfiles de egresados; esta información transmitida a través de las diversas relaciones es esencial en el proceso de obtención de datos.



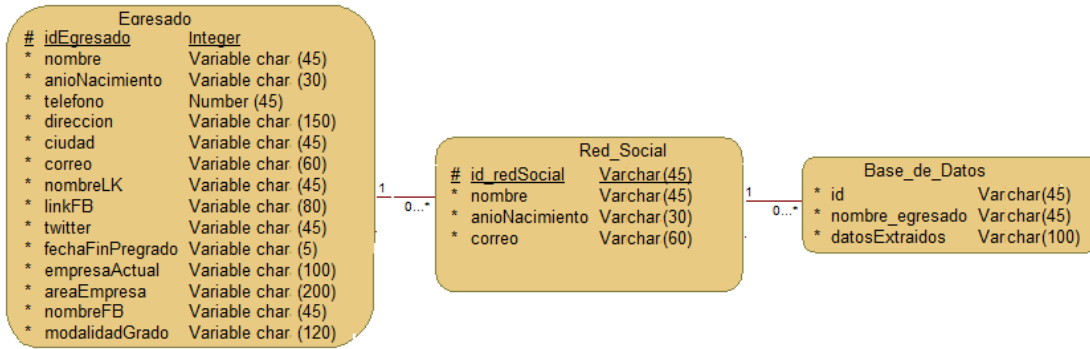


Figura 3. Esquema conceptual de origen de datos

**Esquema físico de las fuentes de datos (Source Physical Schema, SPS):** este esquema define la estructura física de las fuentes de datos que van a alimentar la bodega de datos. Para el caso de la investigación, en la figura 4 se observa cómo se encuentra físicamente la base de datos de las fuentes de datos en un almacén.

Una vez realizado el análisis de las fuentes de datos, se procede a seleccionar las tablas y los respectivos campos que alimentan la bodega de datos, lo cual da una visión clara para el logro del objetivo del proceso de inteligencia de negocios.

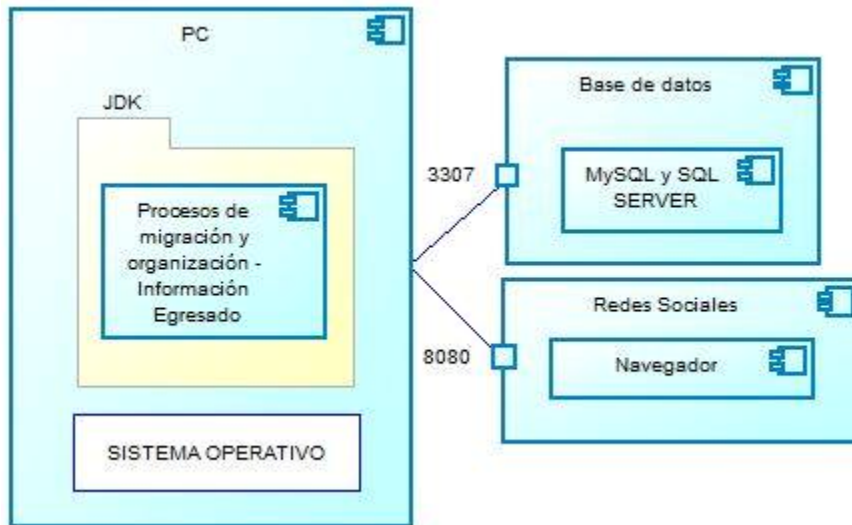


Figura 4. Esquema físico de las fuentes de datos

## Fase de elaboración

En esta fase se busca, por medio de las actividades, extraer y organizar información pertinente al caso de estudio, migrando datos de las redes sociales para que de esta forma se pueda crear una base de datos de una muestra de la población en estudio.

### a. Flujo de diseño

El objetivo principal de este flujo de trabajo consiste en realizar el diseño conceptual de la bodega de datos, además de realizar un primer proceso de selección, limpieza, construcción e integración entre la bodega y las fuentes de datos. Para llevar a término dichas actividades, la metodología BIEP recomienda desarrollar los siguientes esquemas:

**Esquema de estados de máquina en la bodega de datos (*Data Warehouse State Machine Schema, DWMSS*):** este esquema representa el comportamiento dinámico de una entidad con base en su respuesta a los acontecimientos, mostrando la manera en que la entidad reacciona ante los eventos diversos en función de su estado actual (Gerolami, Revello & Venzal, 2011).

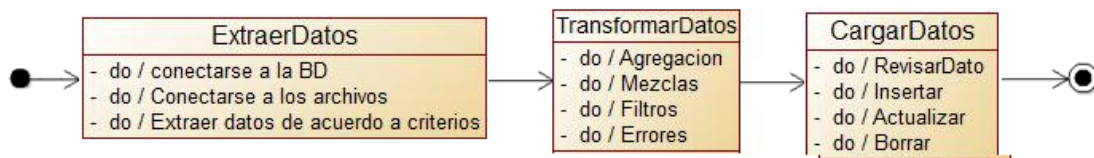


Figura 5. Esquema de estados de máquina en la bodega de datos

En la figura 5 se presentan los estados para hacer el proceso de mapeo del espacio temporal con la tabla *egresado*, pasando por el espacio temporal de la bodega de datos.

**Esquema de actividades de la bodega de datos (*Data Warehouse Activity Schema, DWAS*):** está basado en el diagrama de actividades. Este esquema es el equivalente de los diagramas de flujo de datos en el desarrollo estructurado de la bodega de datos. La figura 6 muestra las actividades que se deben realizar para llevar a cabo la extracción de datos de los perfiles de egresados, poblar la base de datos y, posteriormente, generar un *Data Mart*.

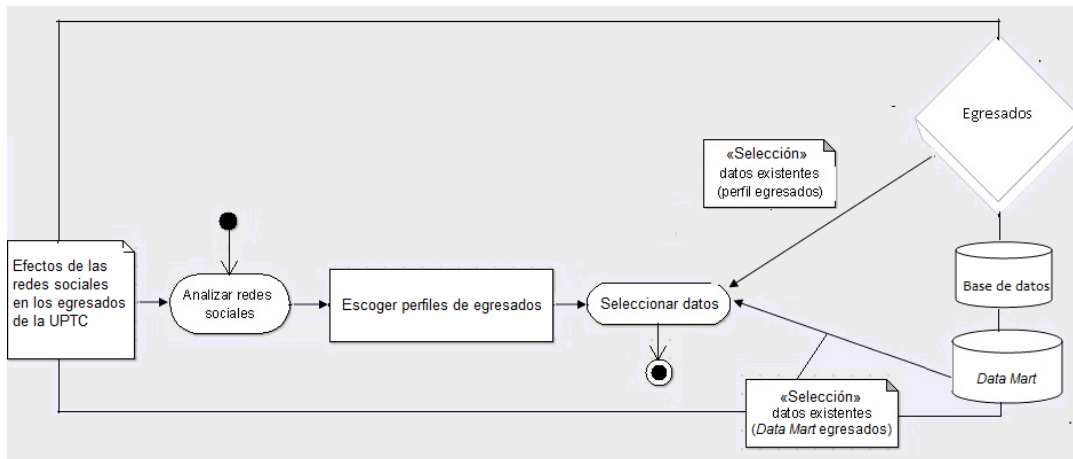


Figura 6. Esquema de actividades de la bodega de datos

## b. Flujo implementación

En este flujo de trabajo, el objetivo principal es crear la bodega de datos y su estructura física para empezar a recibir datos y optimizar el funcionamiento de la misma.

### Fase de construcción

Las actividades que hacen parte de esta fase buscan desarrollar la integración de los datos recolectados con la base de datos de la muestra de egresados para llevar a cabo el diseño de un *Data Mart* (Rico-Sulayes, 2015) con la herramienta SQL Server, con el fin de mostrar adecuadamente la información obtenida de las redes sociales. Para ello, se desarrollan los siguientes flujos de actividades:

#### a. Flujo despliegue

Actividad en la que se determina la forma en que serán utilizados los datos obteniendo la planificación del despliegue y de la monitorización, además de generar el informe final y la revisión del proyecto. Con los datos extraídos, transformados y almacenados en la base de datos, se diseñan consultas que permiten generar reportes estadísticos; estos contribuyen a dar respuesta a los requisitos del negocio planteados en el flujo de requerimientos de la fase de inicio.

#### b. Flujo pruebas

Se pretende verificar que el proceso funcione correctamente, para lo cual se tienen las siguientes actividades:

- Planificar, diseñar e implementar las pruebas mediante la creación de casos de prueba (ver tabla 2).

Tabla 2. Descripción de las pruebas aplicadas

Prueba	Descripción
Análisis de datos	Verificar que la migración se haya realizado limpiamente, es decir, que todos los datos recolectados se encuentren en la base de datos.

Realizar consultas	Elaboración de consultas internas con los elementos de la base de datos para corroborar la información.
Verificar ítem a evaluar	Diseño de consultas que evalúan la experiencia laboral, formación académica e información personal, con el fin de generar reportes.

- Revisar el proceso: teniendo en cuenta el análisis y la verificación, se tienen los aspectos presentados en la tabla 3.

**Tabla 3.** Revisión del proceso de análisis y verificación

Aspecto	Descripción
Influencia de las redes sociales	Se determina la existencia de un porcentaje elevado de egresados interesados en la creación de perfiles en <i>Facebook</i> , <i>Twitter</i> o <i>LinkedIn</i> , así como la constante actualización de sus datos en dichas redes sociales, lo que facilita la extracción de los parámetros establecidos.
Importancia de los datos extraídos	Los datos capturados de las diferentes fuentes de información permiten elaborar reportes estadísticos que facilitan la toma de decisiones por parte de la Escuela de Ingeniería de Sistemas y Computación en cuanto a aspectos como, por ejemplo, la creación de nuevos cursos de posgrado con base en la información existente en el aspecto de áreas de interés y aptitudes de cada egresado.

### Fase de transición

En esta fase final, las actividades planteadas se basan en pruebas necesarias dentro del proceso de migración, estas pruebas se centran en la revisión de la documentación obtenida y en la visualización de información y de oportunidades de mejora dentro del caso de estudio. Para dar cumplimiento se siguen los siguientes flujos:

#### a. Flujo de mantenimiento

- Definir revisiones posdesarrollo que aporten a la adecuada administración de la información:
  - Mantener el contacto con los egresados usando las redes sociales.

- Promover proyectos que involucren a los egresados con las actividades académicas, sociales y culturales de la Escuela.
- Vincular un administrador que realice adecuada gestión de la información con el fin de mantener el contacto con los egresados.
- Incentivar a los estudiantes a propender por la investigación y seguimiento de la temática de egresados, redes sociales e inteligencia de negocios.

## **b. Flujo posdesarrollo**

Para cumplir la finalidad de este flujo de trabajo se plantean las siguientes actividades:

- Revisar la documentación: análisis y evaluación de cada una de las fuentes de datos físicas y digitales que aportan de forma teórica y práctica al desarrollo del proyecto, además de la integración de trabajos de investigación que significan un aporte e innovación a futuras modificaciones.
- Identificar oportunidades de mejora:
  - Promover el desarrollo de nuevos trabajos que aporten a la investigación.
  - Se necesita un administrador que se centre en la actualización de los datos.
  - Elaboración de algoritmos y herramientas que contribuyan a una extracción automática desde las fuentes de datos con el fin de optimizar el proceso.

**84**

## **Resultados**

Las redes sociales son una fuente de datos importante; de allí se obtuvo la mayor parte de la información analizada en el proyecto. Por tal razón, se realizó una caracterización donde se puede evidenciar el uso de las redes sociales por parte de los egresados (figura 7).

La red social Facebook es la más usada por la población de egresados. Sin embargo, esta genera inconvenientes en lo que respecta a la extracción de la información, en tanto muchos de los perfiles estudiados tenían el acceso denegado a los datos por parte de personas que no estuvieran en la lista de amigos. En este punto se optó por estudiar patrones que permitieran identificar correctamente los perfiles en dicha red social, a saber:

- Amigos en común
- Información visible
- Historial de publicaciones y estados
- Fotos
- Lugares

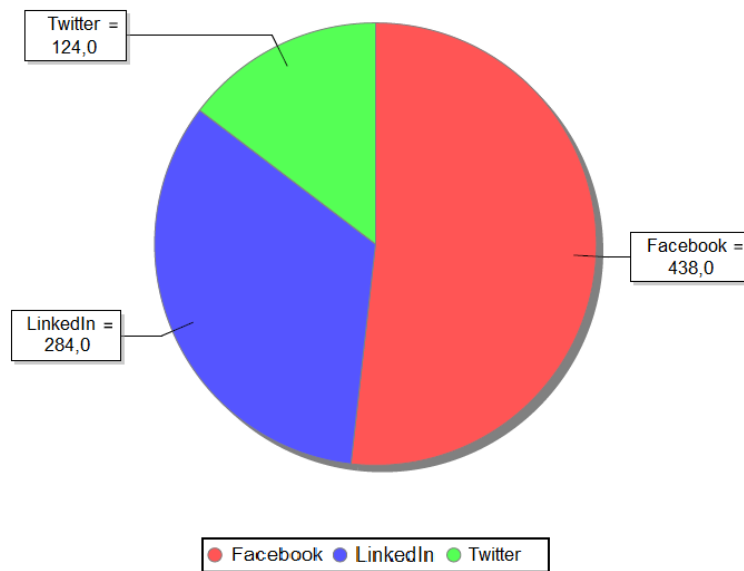


Figura 7. Caracterización del uso de las redes sociales

De la red social *LinkedIn* se pudo percibir que en el *ranking* de las redes sociales estudiadas es la segunda más usada por los egresados. Esta red, que presenta la información de una forma muy organizada —dividiéndola en aspectos personales, laborales y de formación—, fue una de las que más aportó al proceso de extracción. Sin embargo, tiene una pequeña falencia: forma “círculos” según las áreas desempeñadas por cada usuario, lo que limitó un poco el acceso a los datos.

En esta red también se analizaron dos patrones indispensables:

- Los contactos que posee cada persona en el perfil.
- Cómo estaban conectados los perfiles entre sí.

Esta red social forma microrredes internamente según las personas que se tengan agregadas, lo que permite observar la forma como se relacionan las personas que tienen perfiles según las áreas en que se desenvuelven.

Finalmente, Twitter fue la red en la que se encontró la menor cantidad de población de egresados; esta se centra en compartir información que para el caso de estudio no fue de utilidad. En el perfil de cada persona se pueden observar los mensajes compartidos y datos básicos, información que no aportó gran peso para la investigación desarrollada. Por otra parte, al igual que las otras redes, algunos perfiles de Twitter restringen la visualización de su información. El único patrón de estudio que se tuvo en cuenta en esta red social fue las personas a las que sigue cada egresado; por medio de esto, se determinó si el perfil pertenecía a una persona de la población en estudio.

Los datos obtenidos a través de las tres redes sociales nombradas se almacenaron en archivos de Microsoft Excel (XLS) para cargarse posteriormente en una base de datos en MySQL, por medio de una aplicación desarrollada en Java que consta de siete (7) clases:

- *DatabaseLoader*: permite que la aplicación se conecte a MySQL mediante cuatro (4) elementos: *host*, puerto, usuario y contraseña.
- *Egresado*: contiene todas las variables según los datos a cargar en cada tabla.
- *EgresadoData*: contiene cada una de las tablas con las cuales se relaciona la tabla "egresado" (ciudad, país, departamento, areaconocimiento, areaconocimientoegresado, tituloacademico, tituloacademicoegresado y empresa, empresaegresado).
- *Empresa*: es una clase exclusiva para la carga de cada uno de los elementos que hacen parte de la relación débil entre la entidad empresa y el egresado.
- *ForeignTables*: a través de listas agrega los elementos a las tablas que tienen una relación débil con el egresado (empresa, areaconocimiento y tituloacademico)
- *SocialInImporter*: importa los archivos XLS para procesarlos, es la clase principal donde se realiza la ejecución de la aplicación.
- *ReadSocialExcel*: lee los archivos XLS y los envía a la base de datos según el campo y la tabla a la que pertenezcan los datos.

Cuando los datos se cargan en cada una de las tablas de la base de datos, se puede evidenciar sus contenidos a través de MySQL Workbench (figura 8).

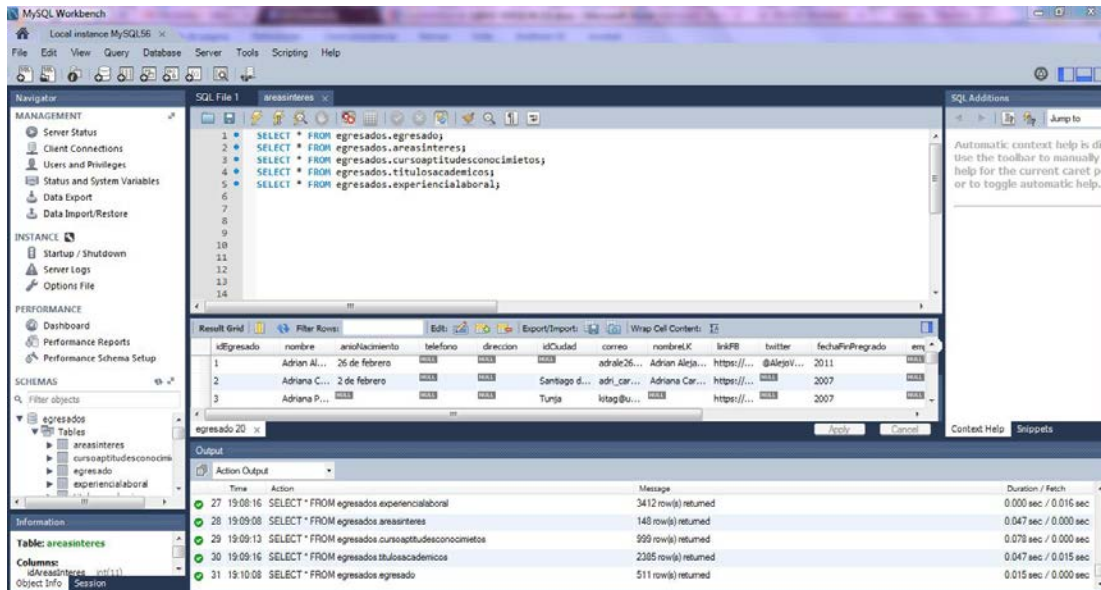


Figura 8. Tablas cargadas en MySQL

## Reportes generados a partir de las consultas

Los reportes que se generaron por medio de la información recolectada fueron estructurados a través de una aplicación en Java donde se ingresaba como parámetro la consulta y esta devolvía un reporte en formato de documento portátil (PDF).

- Egresados por año: la figura 9 muestra la forma en que el programa de Ingeniería de Sistemas y Computación se ha fortalecido con el paso de los años, en tanto tiene cada vez más egresados. Existen periodos de debilidad como 2001, en donde se graduaban las primeras generaciones y la cantidad de egresados era pequeña; sin embargo, de 2007 a 2013 la población de egresados ha crecido, evidenciándose con ello un gusto por la formación en esta disciplina.
- Caracterización de la formación académica de los egresados: los resultados reflejados en la figura 10 muestran que un gran porcentaje de la población posee título de especialización en alguna área en comparación con maestría, doctorado y posdoctorado, ostentados por una población más reducida.



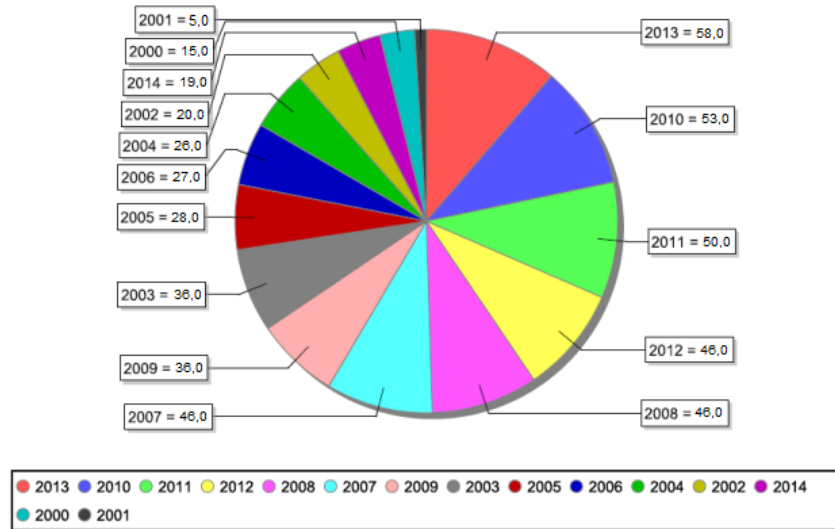


Figura 9. Egresados por año

Es importante observar también que en comparación con la población total estudiada (511 egresados), los porcentajes de personas con posgrados solo llegan a la mitad de la población total, aspecto que debe ser tenido en cuenta para incentivar el estudio de nuevas formaciones que complementen el título de Ingeniería de Sistemas y Computación (figura 10).

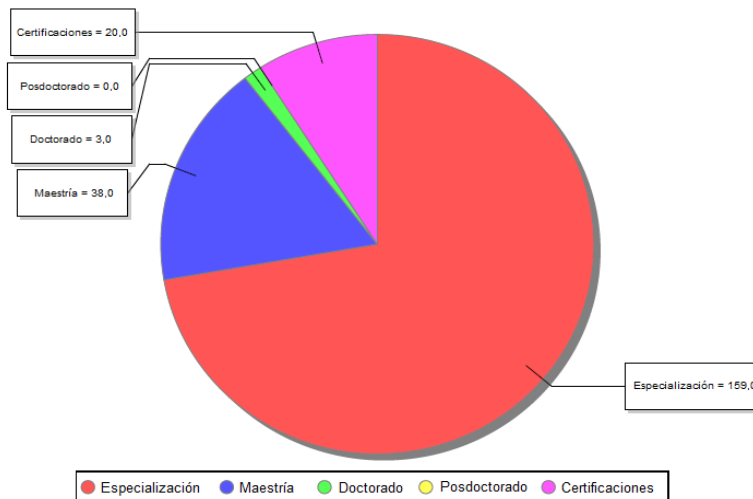


Figura 10. Caracterización de la formación académica de los egresados

- Caracterización modalidad de grado: en la figura 11 se observa que la mayoría de los egresados ha optado por elaborar una tesis como opción de grado. Sin embargo, este aspecto reúne pasantías y participaciones activas en grupos de investigación, pues de aquí se derivan los libros de trabajo de grado que se pueden consultar en la UPTC – seccional Tunja. Por otro lado, desde 2010, cuando se abrió la opción de grado por medio de un semestre de posgrado, se

ha reflejado un aumento significativo de egresados que optan por tomar esta modalidad de grado.

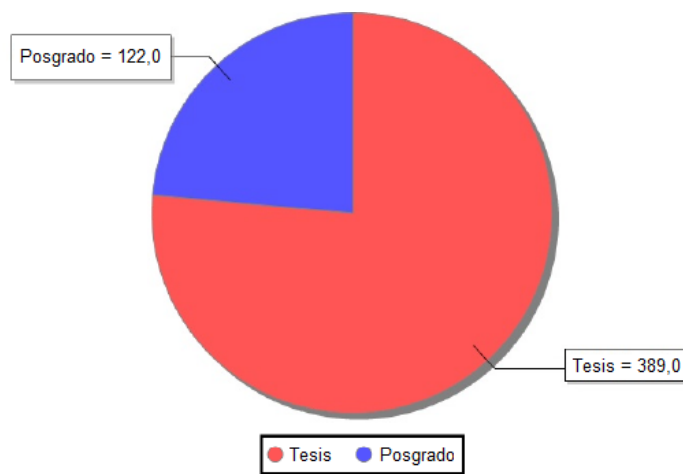


Figura 11. Caracterización modalidad de grado

## Reportes personalizados a través de la aplicación gestión de egresados

Con el ánimo de que fuera posible evidenciar la información de forma clara y ordenada, se desarrolló una aplicación por medio de la cual el usuario puede realizar búsquedas personalizadas según los criterios que desee examinar. La figura 12 muestra la interfaz principal de la misma, donde el usuario puede agregar las condiciones que desea anexar en la búsqueda: el botón "Buscar" arrojará la consulta que se forma a través del criterio de búsqueda y el resultado obtenido; "Guardar resultados" permite conservar estos últimos; y "Nueva búsqueda" permite realizar otras búsquedas con los criterios que se definan para las mismas.

Esta aplicación está diseñada usando como base la técnica de minería de datos de agrupamiento y clasificación (Ballesteros, Sánchez-Guzmán & García, 2013): por medio de asociaciones y ordenaciones de datos que guardan similitud entre sí, se obtienen resultados que para el caso de estudio se orientan a información académica, personal o laboral de la población de egresados.

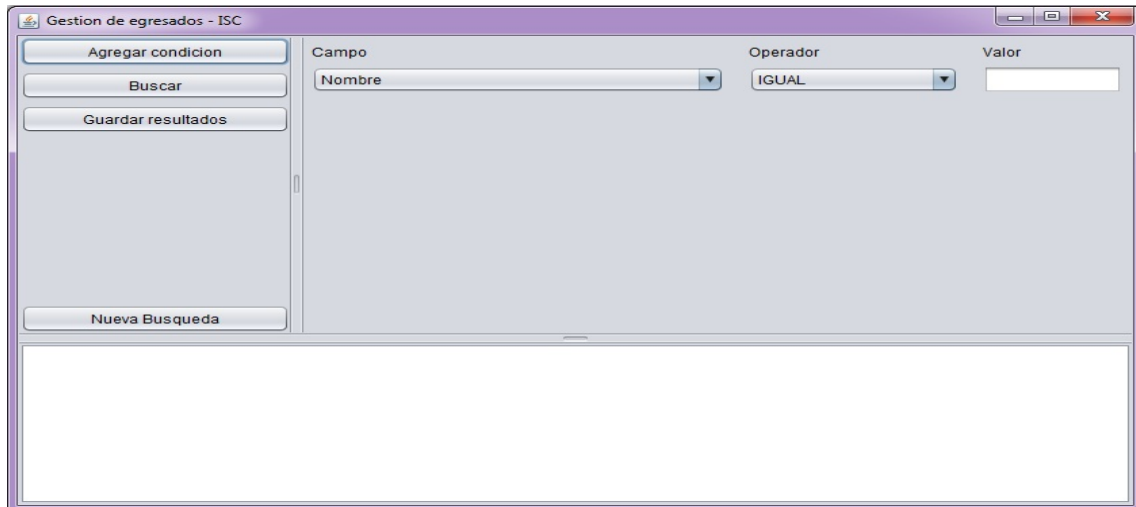


Figura 12. Aplicación de gestión de egresados

## Conclusiones

Las redes sociales estudiadas en la investigación (Facebook, LinkedIn y Twitter) constituyeron las fuentes principales de extracción de datos. Sin embargo, este proceso se complementó también con la aplicación de una encuesta y la interacción directa con los egresados. Aunque las redes sociales son un foco inmenso de información, realizar una gestión de los mismos resulta ser un proceso delicado por cuanto los datos son dinámicos y cambian constantemente.

El manejo de grandes volúmenes de información resulta ser un proceso complejo, más aún cuando los datos son cualitativos. En la investigación se pudo evidenciar la forma en que el agrupamiento y la clasificación de la información se quedan cortos al medir aspectos que describen las áreas personales, laborales y académicas de los egresados.

La creación de una base de datos con la información obtenida de cada egresado facilitó el proceso de selección, organización y detalle de los datos y, a su vez, favoreció al poblamiento de las dimensiones y de los hechos del Data Mart. Sin embargo, en este último aspecto fue difícil realizar algún tipo de medida significativa puesto que los datos relacionados con el tiempo no se encontraban completos.

La incorporación de elementos de inteligencia de negocios permitió la obtención de resultados en los cuales se visualizan reportes generales y personalizados basados en patrones de búsqueda y clasificación según el criterio seleccionado, pero no son relevantes debido al constante cambio que sufre la información a través de las redes sociales.

## Referencias

- Ballesteros Román, A., Sánchez-Gúzman, D. & García, R. (2013). *Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo*. *Lat. Am. J. Phys. Educ.*, 7, 662-668.
- Caldevilla Domínguez, D. (2010). Las Redes Sociales. Tipología, uso y consumo de las redes 2.0 en la sociedad digital actual. *Documentación de las Ciencias de la Información*. Recuperado el 3 de febrero de 2010 en: <http://revistas.ucm.es/index.php/DCIN/article/view/DCIN1010110045A/18656>
- Grau, A. (2009). Tus datos íntimos son una mina. Recuperado de [http://elpais.com/diario/2009/06/03/sociedad/1243980001\\_850215.html](http://elpais.com/diario/2009/06/03/sociedad/1243980001_850215.html).
- Gerolami, N., Revello, E. & Venzal, G., (2011). *Implantación de Data Warehouse Open Free*. (tesis de pregrado). Montevideo: Facultad de Ingeniería, Universidad de la República.
- Gracia del Busto, H., Yanes Enríquez, O. (2012). Bases de NoSQL. *Revista Digital de las tecnologías de la Información y las comunicaciones*, 11, 21-33.
- Herrera Osorio, J. (2011). Metodología Para El Desarrollo De Un Sistema De Inteligencia De Negocios Basada En El Proceso Unificado. (tesis de Maestría en Ingeniería de Sistemas y Computación). Bogotá: Universidad Nacional de Colombia.
- I-Hsien, T., Hui-Ju, W. & Pei-Shan, Chang. (2009). Analyzing Multi-source Social Data for Extracting and Mining Social Networks. Presented at the International Conference on Computational Science and Engineering, Kaohsiung, Taiwan. 815-820.
- Jadav, Jigna & Panchal, Mahesh. (2012). Association Rule Mining Method On OLAP Cube. *International Journal of Engineering Research and Applications (IJERA)*, 2, 1147-1151.
- Martín, A., Chavez, S., Rodríguez, N., Valenzuela, A. & Murazzo, M. (2013). *Bases de Datos NoSQL en Cloud Computing*. Presentado en el XV Workshop de Investigadores en Ciencias de la Computación.
- Rico-Sulayes, A. (2015). Hacia un sistema de ponderación supervisado de bases de datos no estructuradas utilizadas en la construcción de diccionarios especializados. *Revista Facultad de Ingeniería*, 24(38), 97-106.