

Plan de Trabajo Final

Carrera Ingeniería de Sistemas

Facultad de Ciencias Exactas – UNICEN

Tema: Recomendador de productos indexados en un sitio web del tipo e-commerce a usuarios habituales.

Alumno/s: Cristian Ariel Hernández – Valeria Moura Tuya.

Director: Mg. Claudio Aciti.

Codirector: Ing. Luis Federico Bricker

1. Introducción

Un motor de búsqueda, también conocido como buscador, es un sistema informático que busca archivos almacenados en servidores web o en una intranet de una compañía. El primer motor de búsqueda de texto completo fue WebCrawler [1, 2], que apareció en 1994. A diferencia de sus predecesores, éste permitía a sus usuarios una búsqueda por palabras en cualquier página web, lo que llegó a ser un estándar para la gran mayoría de los buscadores [2].

Con el crecimiento de la web empezó a existir una necesidad de proporcionar información procesada a partir de grandes volúmenes de datos que tenían unas estructuras horizontales más o menos similares. Algunas compañías tuvieron que enfrentarse a desafíos con el tratamiento de datos que las tradicionales RDBMS no solucionaban, observando así, que el rendimiento y su velocidad de respuesta eran más importantes que la coherencia, en la que las bases de datos relacionales tradicionales dedicaban una gran cantidad de tiempo de proceso [3]. Es por esta razón que se optó por elegir bases de datos de tipo NoSQL.

NoSQL, también llamado "no sólo SQL", es una amplia clase de sistemas de gestión de bases de datos que difieren del modelo clásico del sistema de gestión de bases de datos relacionales (RDBMS) en aspectos importantes, el más destacado es que no usan SQL como el principal lenguaje de consultas. Los datos almacenados no requieren estructuras fijas como tabla, ni garantizan completamente ACID (atomicidad, coherencia, aislamiento y durabilidad), y están diseñados para manejar la velocidad y escalabilidad de la talla de Google, Facebook, Yahoo, Twitter, entre otros [4].

Los sistemas de recuperación de información son la pieza clave para resolver el problema de la búsqueda rápida y eficiente de información. De hecho, nos encontramos en un momento crucial para estos sistemas. Los más avanzados tratan de agregar información semántica a las búsquedas, con una amplia actividad investigadora en torno a interfaces de usuario, nuevos métodos de búsqueda y nuevas formas de indización, para adelantarse a las necesidades de los usuarios. Hoy en día se debe tener la capacidad de encontrar información y datos de diversos tipos y formatos de manera flexible, libre de forma, que permitan realizar búsquedas requiriendo el mínimo esfuerzo posible [5, 6].

La mayoría de las bases de datos NoSQL fueron construidas con diversos propósitos, uno de ellos en el que se le da uso es el caso del comercio electrónico. El comercio electrónico, también conocido como e-commerce (electronic commerce en inglés), consiste en la compra y venta de productos o de servicios a través de

medios electrónicos, tales como Internet y otras redes informáticas. La cantidad de comercio llevada a cabo electrónicamente ha crecido de manera extraordinaria debido a Internet. Una gran variedad de comercio se realiza de esta manera, estimulando la creación y utilización de innovaciones como la transferencia electrónica de fondos, la administración de cadenas de suministro, el marketing en Internet, el procesamiento de transacciones en línea, el intercambio electrónico de datos, los sistemas de administración del inventario y los sistemas automatizados de recolección de datos [7, 8, 9].

2. Motivación

Las bases de datos relacionales carecen de modelos de datos ricos, consultas dinámicas, y cualquier noción de transaccionalidad. Al mismo tiempo que la velocidad de respuesta en estas últimas es notoriamente más lenta que para el caso de las NoSQL.

Por esta razón es que el uso de bases de datos NoSQL se ajustan muy bien a sitios web dedicados al comercio electrónico, dado que organizar la información en este tipo de bases de datos permite realizar búsquedas dinámicas, podría pensarse en la idea de que cada artículo a vender es un documento y que ese documento tiene detalles del producto. Es entonces que para realizar las búsquedas es necesaria una herramienta eficiente que resuelva consultas personalizadas en poco tiempo.

Se propone realizar una aplicación que recomiende productos basándose en la elección actual del usuario teniendo en cuenta la categorización de los mismos. La organización en categorías se basará en un estudio de mercado, el cual, será implementado utilizando una API de código abierto de recuperación de información mediante la ponderación de los distintos pares <clave,valor>, que en este análisis, representarán a los productos. Las sugerencias de cada recomendación seguirán un orden que se corresponderá con el análisis de mercado que se esté utilizando.

3. Objetivos

Selección y utilización de una API de código abierto para recuperación de información que permita realizar sugerencias ponderadas de productos en un sitio web del tipo e-commerce.

4. Cronograma de actividades

Se propone realizar las siguientes actividades:

- Recopilación bibliográfica necesaria en cada etapa del proyecto.
- Estudio de la bibliografía.
- Selección de una API de código abierto para recuperación de información aplicado al desarrollo de un motor de búsqueda.
- Implementación de aplicación de pruebas.
- Evaluación de resultados / Conclusiones.
- Redacción del informe.
- Presentación de proyecto de grado.

5. Bibliografía

1. Cinto Niqui Espinosa. “La comunicación es vida: Reflexiones eclécticas sobre tics y contenidos audiovisuales”. Editorial UOC. 2011
2. Tom Seymour, Dean Frantsvog, Satheesh Kumar (Minot State University, USA). “ History Of Search Engines, International Journal of Management & Information Systems” – Fourth Quarter 2011, Volume 15, Number 4.
3. Shashank Tiwari. “Professional NoSQL”. John Wiley & Sons. 2011
4. Gaurav Vaish. “Getting Started with NoSQL”. 1st Ed. Packt Publishing Ltd. 2013
5. Marvin Ramírez Valenzo, René E. Cuevas Valencia, José Mario Martínez Castro. “Integration of full-text searches NoSQLdatabases”. Revista Vinculos - Vol 8 – Número 1. 2011.
6. Ricardo Baeza-Yates, Berthier Ribeiro-Neto. “ Modern Information Retrieval”. Addison-Wesley. 1999
7. Editorial Vértice. “E-Commerce: aplicación y desarrollo”. Editorial Vértice. 2010
8. Dotty Boen Oelkers. “Comercio Electrónico”. Cengage Learning Editores. 2003
9. Gary P Schneider. “Comercio Electrónico”. Cengage Learning Editores. 2004

Cristian Ariel Hernández

Valeria Moura Tuya

Avalo la presente solicitud de evaluación,

Mg. Claudio Gustavo Aciti
Director

Ing. Luis Federico Bricker
Co-director