

# **ESPECIALIZACIÓN EN INGENIERÍA DE SOFTWARE**



**UNIVERSIDAD DE LOS LLANOS**

**CURSO: TEXT DATA ANALYTICS**

**MIKE BRIAN OLIVERA  
JAIVER EDREY LESMES MORA**

## Contenido

<b>1. Propuesta de Proyecto .....</b>	<b>2</b>
<b>1.1 Sistema de Recomendación de Libros de Amazon – SIRELA.....</b>	<b>2</b>
<b>1.2 Fuente de datos.....</b>	<b>2</b>
<b>1.3 Los datos .....</b>	<b>2</b>
<b>1.4 Ejemplos de datos .....</b>	<b>3</b>
<b>2. Descripción del Problema.....</b>	<b>3</b>
<b>3. Objetivo y alcance .....</b>	<b>4</b>
<b>3.1 Objetivos general.....</b>	<b>4</b>
<b>3.2 Alcance.....</b>	<b>4</b>
<b>3.3 Diseño preliminar del esquema del sistema propuesto .....</b>	<b>4</b>

# 1. Propuesta de Proyecto

## 1.1 Sistema de Recomendación de Libros de Amazon – SIRELA



El Sistema de Recomendación de Libros de Amazon – SIRELA, proyecto que se realiza para el curso Electiva II - Text Data Analytics, de la Especialización en Ingeniería del Software de la Universidad de los Llanos, el cual tiene el fin de implementar el contenido temático que se trabaja durante el curso, mejorando la capacidad y habilidad en el desarrollo de sistemas de búsqueda o recomendación a través del procesamiento, análisis y modelado de cierto conjunto de datos, seleccionados según la necesidad a satisfacer. El Sistema mejora la experiencia del usuario que realiza sus búsquedas habituales, esto con el fin de realizar una posible compra de libros en la plataforma de venta de libros de Amazon, mejorando la experiencia del usuario, y proporcionalmente mejorando las ventas de la plataforma.

## 1.2 Fuente de datos

El conjunto de datos a ser consumidos y tratados durante el desarrollo y funcionamiento del Sistema, el Book-Crossing Dataset facilitado por Cai-Nicolas Ziegler, del Institut für Informatik, Universität Freiburg de Alemania, el cual es un conjunto de datos de las calificaciones de libros, que se encuentran en la plataforma de Amazon. Estos conjuntos de datos pueden ser consultados en la página web <http://www2.informatik.uni-freiburg.de/~ctiegle/BX/>.

## 1.3 Los datos

Los datos que se desagregan del conjunto de datos mencionado anteriormente, se encuentran almacenados en archivos con extensión .csv, documento de formato abierto en el que se representa su contenido como columnas, las cuales están separadas con comas y las filas por saltos de líneas.

El conjunto de datos se distribuye en tres (3) archivos .csv, los cuales se describen a continuación.

BX-Books: se encuentra compuesto por 271,360 registros, y las columnas ISBN, Título del Libro, Autor del Libro, Año de Publicación, Editorial, enlaces a imagen pequeña, imagen mediana e imagen grande de la portada.

BX-User: compuesto por 278,858 registros, y las siguientes columnas, Código de Usuario, Localización, Edad.

BX-Book-Ratings: documento que lo componen 1'149.780 registros, con las columnas Código de Usuario, ISBN, Calificación del Libro.

## 1.4 Ejemplos de datos

Dentro del archivo BX-Books, las columnas, que contienen las características que tiene los libros, se muestran a continuación, con una fila de información como ejemplo de los datos que lo componen.

ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-M	Image-URL-L
195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	<a href="http://images.amazon.com/images/P/0195153448.01.THUMBZZZ.jpg">http://images.amazon.com/images/P/0195153448.01.THUMBZZZ.jpg</a>	<a href="http://images.amazon.com/images/P/0195153448.01.MZZZZZZZ.jpg">http://images.amazon.com/images/P/0195153448.01.MZZZZZZZ.jpg</a>	<a href="http://images.amazon.com/images/P/0195153448.01.LZZZZZZZ.jpg">http://images.amazon.com/images/P/0195153448.01.LZZZZZZZ.jpg</a>

## 2. Descripción del Problema

La idea del proyecto, ayudada por el interés generado en la temática estudiada en el curso, surge con el ánimo de facilitar y agilizar la manera en que los usuarios realizan la búsqueda de libros según sus preferencias, esto en la plataforma de Amazon en la sección de Libros, la cual está alimentada de una amplia cantidad de información respecto a estos, con una extensa variedad en cuanto a géneros, autores, tiempo en que se lanzaron, entre otros. Al tener toda esa información, la búsqueda de un libro específico puede volverse algo tedioso para el usuario, con una alta posibilidad de que este pierda el interés en seguir buscando libros, se canse de no encontrar algo que le llame la atención, y termine por retirarse de la plataforma sin realizar la elección y compra de un libro de manera exitosa.

### 3. Objetivo y alcance

#### 3.1 Objetivos general

- Desarrollar un sistema que de acuerdo a las calificaciones que tienen los libros publicados en Amazon, permita generar recomendaciones a los usuarios según sus gustos literarios.

#### 3.2 Alcance

Teniendo el conjunto de datos leídos y preparados, se realizan los procesos correspondientes para el modelado, conversión y filtrado de estos con los diferentes métodos necesarios que ayudan en la correcta obtención de los resultados de la búsqueda definida, en procura de que sea lo más efectivo y exacto posible para la comodidad del usuario que utiliza el sistema.

#### 3.3 Diseño preliminar del esquema del sistema propuesto

A continuación, se observa la manera propuesta en que el sistema va a trabajar, con cada una de sus etapas, para generar las recomendaciones resultado de acuerdo al conjunto de datos y la búsqueda.

