

Práctica 2

Recomendador de películas

Objetivo

Por parejas, desarrollar en Python una aplicación original que realice recomendaciones sobre un conjunto de películas, utilizando para ello filtrado colaborativo basado en ítems.

Descripción

Descompondremos el trabajo a realizar en varias etapas.

1. ANÁLISIS DE LA FUENTE DE DATOS

Se trabajará con los datos públicos del dataset de películas y sus valoraciones de la web de MovieLens. Inicialmente con el dataset “small”, caracterizado por:

- Valoraciones: 100.000
- Películas: 9.000
- Usuarios: 600

Puede encontrarse más información [aquí](#), y descargar el fichero con los datos desde [aquí](#). El dataset se encuentra repartido en 4 ficheros: links.csv, movies.csv, ratings.csv y tags.csv.

Deberán analizarse dichos ficheros para entender su significado y relación, para poder avanzar a la siguiente etapa.

2. CARGA DE DATOS

Los datos necesarios deberán transferirse a una **base de datos** creada para dar soporte a la aplicación. Por simplicidad, se sugiere utilizar **SQLite** aunque puede utilizarse cualquier otro SGBD. Deberán crearse las tablas indispensables para representar las diferentes entidades y relaciones necesarias.

Una vez creadas las tablas, deberá realizarse un **proceso automático para transferir los datos de los ficheros CSV a la BD**.

3. APLICACIÓN DE RECOMENDACIONES

El sistema a desarrollar será una aplicación en **Python** con una sencilla **interfaz gráfica** que se conectará a la anterior base de datos. Para un usuario seleccionado, ofrecerá recomendaciones personalizadas mediante el **filtrado colaborativo basado en ítems**. Las **similitudes entre ítems** se calcularán en base al **coseno ajustado**.

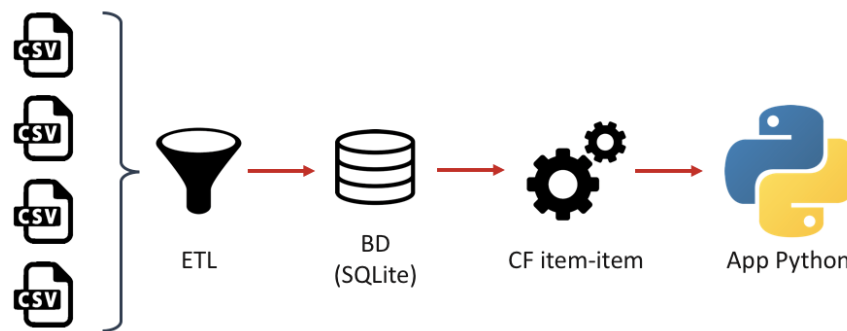


Figura 1. Esquema de la aplicación

Para calcular la predicción será necesario definir previamente el vecindario. Para ello, cada pareja deberá elegir al menos una de estas dos opciones para que el usuario pueda configurarlo:

- Umbral de similitud.
- Número de ítems en el vecindario.

Tras esto el objetivo es poder presentar dos resultados diferentes:

1. **Ranking de las N películas** con la mejor predicción para el usuario U seleccionado. Inicialmente N será igual a 5, pero idealmente debería poderse elegir.
2. **Predicción** para una determinada película elegida por el usuario, seleccionada de entre aquellas que no haya visto ya. Idealmente la película debería poderse localizar a través de un buscador por su título.

Mi Recomendador

Recomendaciones

Selecciona un usuario: ID. 334

Items del ranking

Umbral de similitud:

¡Recomendar!

Ranking:

▼ ID. ítem	▼ Predicción
12	4.6
234	4.4
543	4.4
557	4.3
671	4.2
897	4.1

Selecciona un usuario: ID. 334

Selecciona una película: ID. 481. Peter Pan (Disney)

Predicción: 3.22

¡Predecir!

Figura 2. Sugerencia de interfaz

Entregables y documentación

La entrega consistirá en un único archivo comprimido en **formato ZIP** (no se aceptan entregas en RAR, 7Z u otros formatos).

Seguir las siguientes indicaciones para adjuntar el envío a través del Campus Virtual. No cumplir estas normas supondrá la no aceptación de la entrega:

1. Crear una **carpeta** cuyo nombre seguirá el siguiente formato: Grupo_**N**.SSII.P2, donde N es el número del grupo de trabajo.
2. Crear una carpeta “documentación”, que contendrá:
 - a. Fichero Grupo_**N**.SSII.**P2**.PDF con todos los detalles de la actividad.
 - b. Incluirá, al menos, los siguiente apartados:
 - Portada
 - Índice
 - Descripción de la BD: diagrama, tablas, atributos, tipos de datos, etc.
 - Descripción del proceso de carga de datos a la BD: CSV → BD
 - Recomendaciones:
 - Descripción y justificación de las funciones/métodos implementados, junto con sus pseudocódigos.
 - Resultados y conclusiones
 - Manual de instalación y dependencias.
 - Instrucciones para instalar la base de datos.
 - Manual de usuario
 - Bibliografía
3. Crear una carpeta “código” que incluya:
 - a. Código fuente de la aplicación, junto con los diferentes ficheros necesarios para su ejecución (CSVs o fichero de la BD en su caso, etc.)
4. Comprimir en **formato ZIP** la carpeta raíz contenedora de todos los ficheros, especificada en el punto 1. El nombre deberá seguir el siguiente formato: Grupo_**N**.SSII.P2.**ZIP**

Deberá subirse al Campus Virtual a través de la actividad correspondiente **antes de las 23:55h** de la fecha marcada como límite.