

Proyecto Webir 2015

MovieRecommender

Grupo 3

Federico González 4.587.392-1

Matilde Macciò 3.245.946-9

Contenido

Introducción	3
Implementación.....	3
Tecnologías utilizadas	3
Arquitectura y desarrollo de la aplicación	4
Migración de datos.....	4
Aplicación Web	5
Librería NReco.Recommender ^[5]	5
Caso de Prueba	6
Conclusiones y trabajos futuros	7
Referencias.....	7

Introducción

El proyecto trata sobre un sitio de películas online, el cual genera recomendaciones a usuarios basado en sus preferencias y en las preferencias de usuarios similares en gustos. También permite aplicar filtros sobre estas recomendaciones y realizar búsquedas por más de un filtro, obteniendo así un listado de resultados con información más completa de las películas. El usuario puede puntuar las mismas generando preferencias. Se provee de la funcionalidad de ver la película online embebida en el propio sitio.

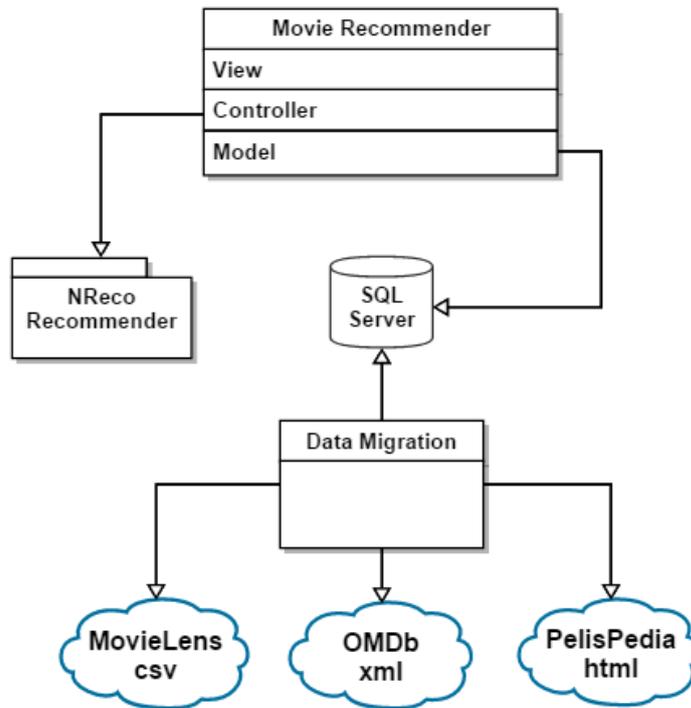
Para abordar este proyecto se debió generar una base de datos de películas, usuarios y ratings de películas por usuarios. Para esto se obtuvo la información de diferentes sitios generando una base estática. Además se utilizó una librería para resolver las recomendaciones a los usuarios (filtrado colaborativo).

Implementación

Tecnologías utilizadas

La aplicación fue desarrollada en ASP .Net Framework 4.5.
Los lenguajes utilizados fueron C#, jQuery y HTML 5.
Como motor de base de datos, SQL Server Express Edition.
La librería para filtrado colaborativo, NReco.Recommender.

Arquitectura y desarrollo de la aplicación



El sistema cuenta con dos módulos: una aplicación de consola para la migración de datos y la aplicación web.

Migración de datos

Del sitio MovieLens ^[1] se obtuvo un juego de datos reducido en formato .csv. Del mismo se obtuvieron listados de películas, usuarios y ratings.

Luego, consultando la API OMDb ^[2] se obtuvieron más datos sobre cada película, como trama, actores, directores, rating en IMDb ^[3], etc.

Dicha consulta tiene la siguiente forma:

<http://www.omdbapi.com/?i=tt0114709&plot=full&r=xml>

Donde:

- tt0114709 es el identificador en IMDb de la película dado por el juego de datos de MovieLens.
- plot=full solicita la trama completa de la película
- r=xml es el formato en el que se devuelven los datos

El siguiente es un ejemplo de respuesta de esta API:

```
<?xml version="1.0" encoding="UTF-8"?><root response="True"><movie title="Toy Story" year="1995" rated="G" released="22 Nov 1995" runtime="81 min" genre="Animation, Adventure, Comedy" director="John Lasseter" writer="John Lasseter (original story by), Pete Docter (original story by), Andrew Stanton (original story by), Joe Ranft (original story by), Joss Whedon (screenplay), Andrew Stanton (screenplay), Joel Cohen (screenplay), Alec Sokolow (screenplay)" actors="Tom Hanks, Tim Allen, Don Rickles, Jim Varney" plot="A little boy named Andy loves to be in his room, playing with his toys, especially his doll named &quot;Woody&quot;. But, what do the toys do when Andy is not with them, they come to life. Woody believes that he has life (as a toy) good. However, he must worry about Andy's family moving, and what Woody does not know is about Andy's birthday party. Woody does not realize that Andy's mother gave him an action figure known as Buzz Lightyear, who does not believe that he is a toy, and quickly becomes Andy's new favorite toy. Woody, who is now consumed with jealousy, tries to get rid of Buzz. Then, both Woody and Buzz are now lost. They must find a way to get back to Andy before he moves without them, but they will have to pass through a ruthless toy killer, Sid Phillips." language="English" country="USA" awards="Nominated for 3 Oscars. Another 22 wins & 16 nominations." poster="http://ia.media-imdb.com/images/M/MV5BMTgwMj4MzU5N15BMl5BanBnXkFtZTcwMTMyNTk3OA@@_V1_SX300.jpg" metacore="92" imdbRating="8.3" imdbVotes="565,387" imdbID="tt0114709" type="movie"/></root>
```

Por último, para obtener los videos online a embeber en el sitio, se aplicó scraping sobre PelisPedia^[4]. Debido a que los títulos de las películas en PelisPedia no siempre coinciden con los generados en la base de datos de la aplicación, ya sea porque se encuentran en otro idioma o porque contienen más información, se debió realizar primero una búsqueda por título de esta forma:

<http://www.pelispedia.tv/buscar/?s=movieTitle>

obteniendo como resultado un código html con una lista de posibles candidatos. Luego sobre esta lista se comparó cada elemento por el título exacto en un tag determinado. Una vez obtenido el elemento, se obtuvo la url exacta de la película, y de ahí el iframe conteniendo el video online. El código de esta iframe fue guardado en la base para cada película.

Aplicación Web

La arquitectura de esta aplicación sigue el patrón MVC (model, view, controller). Donde la capa de modelo se comunica con la base de datos para obtener información y persistir usuarios y preferencias. La capa de vistas es la que implementa la interfaz del usuario y contiene código jQuery que permite embeber los videos y manejar de forma dinámica los datos a través del controlador. La capa del controlador es la que tiene la lógica de la aplicación, y se comunica con la .dll de NReco.Recommender para obtener recomendaciones para el usuario logueado.

Librería NReco.Recommender^[5]

Este es un motor de filtrado colaborativo que basándose en las preferencias del usuario, genera un vecindario (de tamaño configurable) de usuarios con preferencias similares. Donde usuarios similares son los que comparten un conjunto de preferencias. Retornando un conjunto de películas para cada una de las cuales asigna un peso que indica la probabilidad de que al usuario le guste la película. La cantidad de películas a recomendar también es configurable.

Caso de Prueba

A continuación se muestra una recomendación obtenida para un usuario que comparte la mayoría de las preferencias con otro.

El usuario user_1 tiene las siguientes preferencias:

	userName	title	rating
1	user_1	Blade Runner	1
2	user_1	Fargo	3
3	user_1	Memento	4
4	user_1	Shutter Island	4
5	user_1	Inception	5
6	user_1	Inside Out	5

Y el usuario user_2:

	userName	title	rating
1	user_2	Blade Runner	2
2	user_2	Fargo	3
3	user_2	Memento	5
4	user_2	Shutter Island	5
5	user_2	Inception	5

Al pedir recomendaciones para el usuario user_2, se obtiene como resultado la película Inside Out, que forma parte de las preferencias de user_1 pero aún no fue puntuada por user_2.

The screenshot shows a web application titled "Movie Recommender" with a user profile "user_2" in the top right. The search interface includes input fields for Title, Actor, Director, Writer, and Year, along with a Genre dropdown menu and "Search" and "Recommend" buttons. Below the search area, a recommendation for the movie "Inside Out" is displayed. The recommendation includes a movie poster, an IMDb Rating of 8.4, a year of 2015, a runtime of 94 minutes, genres of Animation and Comedy, directors Pete Docter and Ronnie Del Carmen, and actors Amy Poehler, Phyllis Smith, Richard Kind, and Bill Hader. A plot summary is provided, and there is an "Open video" button.

userName	title	rating
user_1	Blade Runner	1
user_1	Fargo	3
user_1	Memento	4
user_1	Shutter Island	4
user_1	Inception	5
user_1	Inside Out	5
user_2	Blade Runner	2
user_2	Fargo	3
user_2	Memento	5
user_2	Shutter Island	5
user_2	Inception	5

Conclusiones y trabajos futuros

Existen muchos sitios con información de películas, en distintos idiomas, donde algunos permiten ver las películas online y/o descargarlas, otros sólo proveen información y puntuación sobre las mismas. Esto hace que la forma de obtener los datos sea muy variada.

Para este prototipo nos enfocamos en una base reducida pero que contuviera información lo más completa posible, que se pudiera ver videos online embebidos en el sitio, y que contara con usuarios y sus preferencias para poder así aplicar filtrado colaborativo. Para ello fue necesario integrar información de tres fuentes diferentes.

La herramienta de filtrado colaborativo funcionó como una caja negra. Debimos implementar una interfaz que permitió la comunicación entre la librería (NReco.Recommender) y nuestra base de datos.

Como trabajo a futuro, sería interesante obtener información de películas de forma dinámica, y de mayor cantidad de fuentes. Sería bueno generar un diccionario para manejar películas en diferentes idiomas. También se podría integrar la aplicación web con redes sociales, y poder compartir recomendaciones a otros usuarios. Para bases de datos mayores sería conveniente agregar índices Full Text que optimizan la búsqueda en catálogos sobre textos.

Referencias

- [1] <http://grouplens.org/datasets/movielens/>
- [2] <http://www.omdbapi.com/>
- [3] <http://www.imdb.com/>
- [4] <http://www.pelispedia.tv/>
- [5] http://www.nrecosite.com/recommender_net.aspx