Rafael Gonzalez Villarreal y Julio de la Olla (Segundo apellido)

  Ingeniería Informática del Software

Trabajo AII: Peliculas

**Introducción**

Como trabajo propuesto para la asignatura de Acceso Inteligente a la Información hemos pensado en el desarrollo de un sistema que realice en primer lugar la extracción de la página de Sensacine (concretamente del siguiente enlace: <http://www.sensacine.com/peliculas/mejores/nota-espectadores/> ).

Una vez que realizamos la extracción de los datos los almacenamos en un csv. A la vez y para usar la tecnología de Whoosh creamos también los índices y guardamos la información sobre cada película.

Finalmente para poder mostrar una interfaz atractiva y a la vez usar otra tecnología más de las aprendidas en clase, hacemos uso de Django. Gracias a Django aparte de tener una interfaz más atractiva, gracias a bootstrap, que la que se puede tener con Tkinter. En este sistema creado se podrá realizar tanto búsqueda de películas mediante Django o Whoosh.

**Tecnologías**

Para el desarrollo del trabajo hemos usado las siguientes tecnologías aprendidas en clase:

|  |  |
| --- | --- |
| Recogida de información de web | Beautyfulsoup |
| Búsqueda y almacenamiento mediante índices | Whoosh |
| Interfaz y búsqueda | Django |
|  |  |

**Desarrollo del trabajo:**

Para comenzar el trabajo necesitamos que página usaríamos para la extracción de los datos, en primer lugar pensamos en la página de Mk2 pero debido a la dificultad de navegar por ella y encontrar algo claro que extraer decidimos usar otra conocida página para la extracción de los datos, Sensacine, concrétame del enlace colocado en la introducción de este documento.

Una vez observada la página comenzamos a escribir el código Python de Beautyfulsoup para extraer los datos de cada dato que consideramos importante adquirir: Titulo de la película, directores, reparto, fecha de estreno o re-estreno, genero, sinopsis y los votos de los medios, usuarios y sensacine.

Una vez extraído los datos se realiza el almacenado por índice para el uso de Whoosh en la búsqueda y por otro lado se realiza el almacenado de los datos en un csv que posteriormente se usa con Django para una búsqueda también.

Una vez creado todo el apartado de extracción de datos procedimos a crear los modelos de Django para poder realizar el populate. Cuando conseguimos que el populate funcionase correctamente comenzamos a pensar que acciones podríamos realizar con esos datos. Desarrollamos en primer la búsqueda de película tanto usando Whoosh como otra opción que hacia la búsqueda en Django.