UNIVERSIDAD DE GRANADA

Recuperación de Información

Práctica 3

Implementación de un Sistema de Recuperación de Información utilizando Lucene

Diseño, indexación, búsqueda e indexación de facetas

David Carrasco Chicharro
Daniel Terol Guerrero

1. Indexación

Para poder realizar el proceso de indexación se ha creado una clase, *ExtraerPelícula*, que almacena la información relacionada a una película. A partir del archivo *wiki_movie_plots_deduped.csv*, se ha leído el fichero entrada a entrada, extrayendo los diferentes campos a través de una biblioteca capaz de manejar archivos CSV de forma cómoda y eficiente.

Cada entrada del archivo CSV, se ha almacenando en un vector de objetos de la clase *ExtraerPelícula* para, posteriormente, ser procesado de forma completa todo el *dataset* y realizar así la indexación:

Se recorre el vector de películas y, por cada película, se establecen los diferentes campos de indexación.

a. Título

b. Director

c. Reparto

d. Género

e. Origen

Además de ser campos de indexación, *Género* y *Origen* se han establecido como *Facetas*, pudiendo así, posteriormente, filtrar los resultados de la búsqueda a partir de estos campos.

El índice creado genera varios ficheros que se almacenan en el directorio */indice*.

2. Búsqueda

Para realizar la búsqueda, se hace uso de una interfaz donde se puede establecer el campo de búsqueda deseado y la consulta. Una vez se introduce la consulta, se realiza una *Query* a través de un *parser*.

```
1 Query q = parser.parse(campo+":\""+consulta+"\"");
```

Es entonces, a partir de la consulta y del número de documentos que se quieren mostrar como resultado, cuando se realiza la búsqueda en el índice.

```
TopDocs tdc = FacetsCollector.search(searcher, q, num_docs, fc);
```

A partir del resultado se recorren todos los documentos incluidos en él y se muestra el título de los diferentes documentos.

```
String str = "";
for (ScoreDoc sd : tdc.scoreDocs) {
    Document d = searcher.doc(sd.doc);
    str+=sd.score + "\t\r-␣␣␣"+d.get("titulo")+"\n";
}

jTextAreaPelisEncontradas.setText(str);
jTextAreaPelisEncontradas.update(jTextAreaPelisEncontradas
.getGraphics());
```

Una vez se tienen los resultados, aparecen las categorías asociadas a la búsqueda.

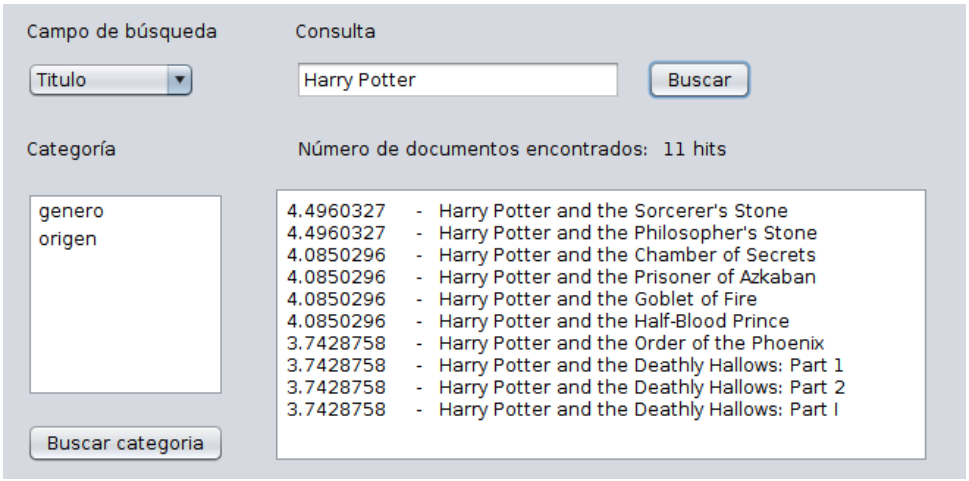

Figura 1: Resultado de búsqueda

Al seleccionar una de las categorías se muestran las diferentes facetas asociadas a la consulta, pudiendo seleccionar una faceta y, por tanto, filtrando la búsqueda a través de una categoría y faceta determinada.

Para poder realizar el filtrado, se hace uso de *DrillDownQuery* y *DrillSideways*.

```
DrillDownQuery dq = new DrillDownQuery(fconfig, q));
   [...]
DrillSideways ds = new DrillSideways(searcher, fconfig,
   taxoReader);
DrillSideways.DrillSidewaysResult dsresult = ds.search(dq,10);
```

Al filtrar por categorías y faceta, el resultado queda como se muestra a continuación:

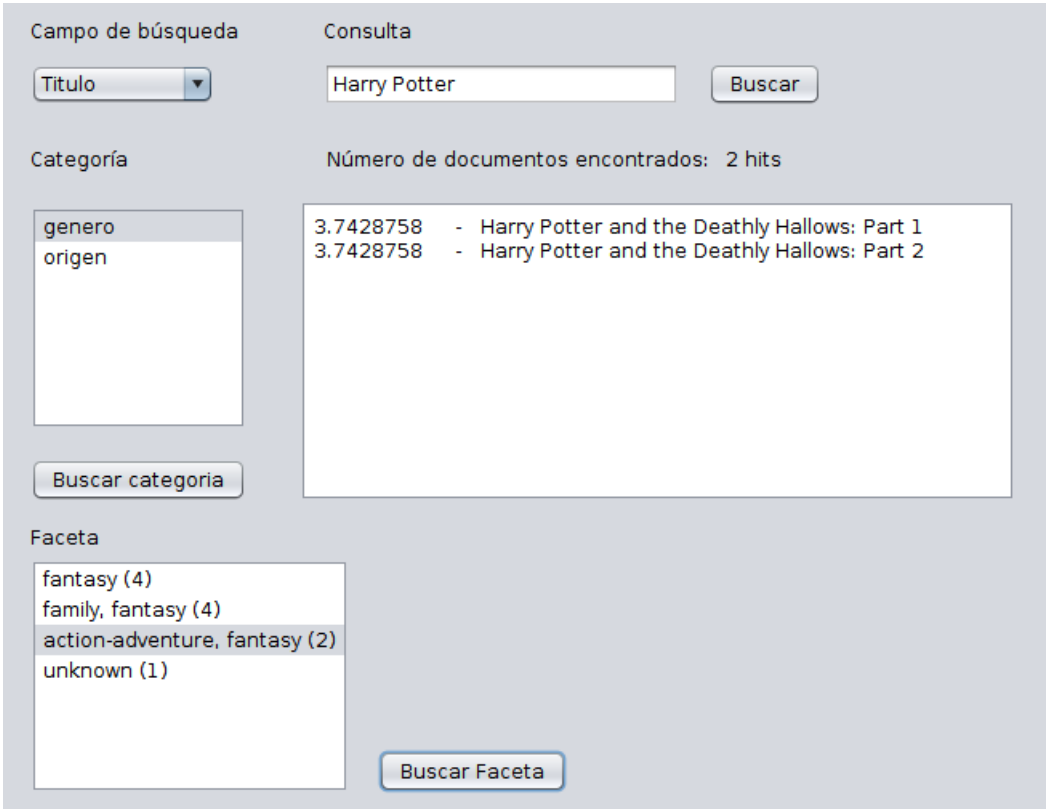

Figura 2: Resultado de búsqueda por facetas

3. Ejecución

Para ejecutar el programa hay que añadir el archivo CSV *wiki_movie_plots_deduped.csv* al directorio del proyecto y ejecutar el archivo *ejecucion.sh*.