

Práctica 4

Búsqueda de información mediante la librería Lucene

Marta Blanco Jaime – 562526
Óscar FracaFerrández – 566416
Alberto Gómez Climente - 698683

ÍNDICE

TAREA 1: INTRODUCCIÓN A LA LIBRERÍA LUCENE	3
RESULTADOS.....	3
CONCLUSIONES	4
TAREA 2: INDEXACIÓN Y BÚSQUEDA DE DOCUMENTOS DE UN DETERMINADO DIRECTORIO.....	4
1ª APLICACIÓN	4
2ª APLICACIÓN	4
METODOLOGÍA	4
HERRAMIENTAS UTILIZADAS	5
DIFICULTADES ENCONTRADAS	5

Tarea 1: Introducción a la librería Lucene

Para obtener los resultados de las consultas pedidas, se ha modificado el método **main** para que sean introducidas por pantalla. El resultado de la modificación es el siguiente:

```
Introduzca query: [FIN para salir] tomcat
Introduzca query: [FIN para salir]tomcat OR lucene
Introduzca query: [FIN para salir]"Lucene"
Introduzca query: [FIN para salir]of
Introduzca query: [FIN para salir]tomcat*
Introduzca query: [FIN para salir]jakarta
Introduzca query: [FIN para salir]Lucene --apache
```

Además, se han cambiado los mensajes por pantalla para que se vea en todo momento a que consulta se hace referencia al arrojar los resultados.

Resultados

Resultado obtenidos en la ejecución con la clase **StandardAnalyzer**:

```
Query: tomcat, found 1 hits.
1. .\ficheros/dos.txt 0.54986566
Query: tomcat OR lucene, found 3 hits.
1. .\ficheros/dos.txt 0.21883568
2. .\ficheros/uno.txt 0.12118581
3. .\ficheros/tres.txt 0.05404176
Query: "Lucene", found 2 hits.
1. .\ficheros/uno.txt 0.4003836
2. .\ficheros/tres.txt 0.17854758
Query: of, found 0 hits.
Query: tomcat*, found 1 hits.
1. .\ficheros/dos.txt 1.0
Query: jakarta, found 0 hits.
Query: Lucene --apache, found 0 hits.
```

Resultado obtenidos en la ejecución con la clase **SimpleAnalyzer**.

```
Query: tomcat, found 2 hits.
1. .\ficheros/dos.txt 0.64329314
2. .\ficheros/dos.txt 0.5360776
Query: tomcat OR lucene, found 4 hits.
1. .\ficheros/dos.txt 0.25829148
2. .\ficheros/dos.txt 0.21524292
3. .\ficheros/uno.txt 0.1361942
4. .\ficheros/uno.txt 0.11349517
Query: "Lucene", found 4 hits.
1. .\ficheros/uno.txt 0.4570735
2. .\ficheros/uno.txt 0.38089457
3. .\ficheros/tres.txt 0.20382795
4. .\ficheros/tres.txt 0.19044729
Query: of, found 4 hits.
1. .\ficheros/dos.txt 0.30384874
2. .\ficheros/cuatro.txt 0.19489677
3. .\ficheros/tres.txt 0.17698653
4. .\ficheros/uno.txt 0.16241398
Query: tomcat*, found 2 hits.
1. .\ficheros/dos.txt 1.0
2. .\ficheros/dos.txt 1.0
Query: jakarta, found 0 hits.
Query: Lucene --apache, found 0 hits.
```

Conclusiones

Las diferencias que se aprecian en la ejecución con una clase y con la otra son debidas a las propias características de cada una. Con **StandardAnalyzer** las palabras comunes son ignoradas y no computan (stopwords), como es el caso de **of** o **OR**. Sin embargo, estas palabras si son consideradas con **SimpleAnalyzer** y las incluye en los resultados encontrados.

Tarea 2: Indexación y búsqueda de documentos de un determinado directorio

Para que las palabras a encontrar se consideren en español, se añadido la clase **SpanishAnalyzer**, y para el tamaño del directorio, en lugar de **RAM** se hace uso de **FSD**.

1ª Aplicación

En esta ocasión, se han vuelto a hacer modificaciones en el código para introducir también por pantalla el directorio en el que se deben buscar los ficheros **“.txt”**, en concreto con el método **recorrerDirectorio**. Además, si este tiene a su vez subdirectorios, también los recorridos.

De esta manera, se crean el índice en la misma carpeta seleccionada por el usuario.

2ª Aplicación

No se ha hecho modificaciones especiales para esta segunda versión ya que desde un principio se ha modificado el código para introducir manualmente las consultas.

En concreto, la modificación que se hizo fue que las palabras claves introducidas por el usuario se van guardando en un array para posteriormente recorrerlo y buscar las en los ficheros encontrados en el directorio.

Metodología

La primera parte de la práctica se resolvió entre los tres miembros fuera del horario de las propias prácticas. La segunda parte, debido al retraso en la realización de la práctica anterior, se decidió que Marta resolviera las modificaciones en el código una vez decidido entre todos que se debería hacer, ya que es la que más manejo de Java tiene.

En esta ocasión, no ha supuesto mucho tiempo de realización ya que se ha obviado la parte optativa por el retraso comentado anteriormente, y las horas dedicadas han sido

	Tarea 1	Tarea 2	Memoria	TOTAL
Alberto	2h	1h	1h	4h
Marta	2h	3h	1h	6h
Oscar	2h	1h	1h	4h

Herramientas utilizadas

Para la implementación del código y ejecución del mismo se hizo utilizando Eclipse.

En cuanto a la comunicación entre los miembros del grupo, esta se ha hecho a través de Whatsapp y GitHub.

Dificultades encontradas

La única dificultad encontrada en esta práctica fue la de averiguar cómo incluir las librerías de Lucene en Eclipse, ya que en un primer momento, al cargar el proyecto, no las encontraba y por tanto no reconocía la mitad del código.