

# SIW - P6 - SEARCH

## Primera Parte

Se han realizado ajustes en las clases BagOfWords y Indexer para adecuarlos a los nuevos tests.

Para las pruebas de search, se ha creado, en la clase Indexer, un método llamado “search”. Este método tiene como objeto devolver una lista ordenada, con un límite de tamaño, de la similitud coseno para da documento que está indexado.

Para realizarlo, el método “search” utiliza varios métodos auxiliares:

- create\_vector\_query
- create\_vector\_\_documentX
- dot\_product
- mod\_documentx
- cosine\_similarity

Todos estos métodos se encuentran debidamente documentados en el código.

He tenido problemas a la hora de ordenar la lista de final, la solución que he encontrado ha sido esta:

```
result = sorted(sorted(result, key=lambda x: x[0], reverse=True), key=lambda x: x[1], reverse=True)
```

Es decir, ordenar la lista además de por el valor, por el texto.

## Segunda Parte

Por último he creado el método “read\_cran” que recibe los nombres de los ficheros que contiene los textos y las queries, y el límite de queries que se probarán (tendrá 5 de valor por defecto).

Para ejecutarlo se debe introducir por parámetro los dos ficheros y, de forma opcional, el límite.

La función de este método es, por tanto, la de buscar las queries de un fichero en los textos de otro.