

SIW - P6 - SEARCH

Se han realizado ajustes en las clases BagOfWords y Indexer para que los tests de ambas clases pasarán.

Para las pruebas de search, se ha creado un método llamado “search” que llama a otro llamado “cosine_similarity”, auxiliar del primero.

El método auxiliar “cosine_similarity” se encarga de calcular las diferentes similitudes del coseno, devolviendo pares de [texto, similitud_coseno]. Esta similitud se calcula para cada documento con la query que se introduce en el método “search”.

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.

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.