

# SIW - P5 - INDEXER

## indexer\_v1

En el fichero `indexer_v1.py` se ha implementado la clase `Indexer` y la clase `BagOfWords` (copiada de prácticas anteriores, con algunas modificaciones).

En lo que a la clase `Indexer` respecta, se han implementado los métodos principales:

- `index`
- `score`
- `dump`

Sobre estos métodos no caben muchas aclaraciones más que en el caso de `score`, por cumplir lo establecido en los test (usando todos los parámetros), se crea con el texto recibido un bag of words y se hace el score de todo él.

Además de los métodos mencionados se han creado los métodos auxiliares: `getTF` y `getIDF`, que calculan el TF y el IDF, respectivamente, con los datos introducidos por parámetro y los propios atributos del objeto `Indexer`.

## txt\_to\_json

En el fichero `txt_to_json.py` implementamos los métodos necesarios para, dado un fichero, transformar este, mediante bag of words e `indexer` en un fichero json.

Además se incluye un método para cargar el fichero recién creado, en el se imprime datos relacionados con este fichero una vez obtenido el fichero.

Los métodos se encuentran documentados en el código.

Para probarlo se ha usado el fichero `cran-1400.txt` que se encuentra en la carpeta del proyecto.