

Documentation du code

Preprocessing.py :

Récupération des données

Correction des accents

Suppression des stop-words

Tokenisation

Tf_idf.py

Calcul les scores pour chaque document en utilisant le système tf-idf

La recherche booléenne est prise en compte aussi

Bm25_system.py

Calcul les scores pour chaque document en utilisant le système bm25

Semantic_search.py

Calcul les scores pour chaque document en utilisant le modèle SBERT pour la recherche sémantique

Global_système.py

Permet de faire le classement général en utilisant un système de pondération

$0.5 * \text{bm25} + 0.4 * \text{SBERT} + 0.1 * \text{tf-idf}$

Paths.py

Pour gérer les différents fichiers

Requirements.txt

Pour avoir la liste de toutes les librairies utilisées

Interface.py

Pour gérer tout ce qui concerne l'interface (correction orthographique, affichage ...)