

## Description du projet :

L'objectif de ce projet est de développer un système d'indexation et de recherche qui permet de stocker, indexer, et rechercher efficacement des informations dans un corpus de documents textuels.

### Objectifs principaux :

1. **Indexation des données**
2. **Recherche de données**
3. **Interface utilisateur**

### Étapes de réalisation :

1. **Analyse et préparation des données :**
  - Choisissez un jeu de données publique
  - Analyse des données
2. **Construction de l'index :**
  - Parcourir le corpus pour extraire les mots et leurs positions.
  - Organiser les données dans une structure comme un dictionnaire.
3. **Implémentation de la recherche :**
  - Créer des algorithmes pour rechercher des mots ou expressions.
  - Afficher les résultats avec des informations telles que le document et la position des termes.
4. **Documentation et interface :**
  - Documenter le code et expliquer le fonctionnement du système.
  - Créer une interface utilisateur conviviale qui permet :
    1. Insérer un mot ou une expression
    2. Appliquer une recherche dans tous les documents
    3. Afficher une carte d'information liée à l'entrée :
      - Nombre d'apparition dans l'ensemble des documents
      - Distribution des indices du mot au sein des différents documents
      - La moyenne d'apparition de l'expression dans chaque document (uniquement quand il est présent)
      - Afficher un extrait de l'expression au sein d'un texte
      - Afficher un word cloud des mots (mot+1 et mot-1) qui entourent l'expression

### Livrables :

- Notebook bien documenté
- Script de l'interface

**Instructions :**

Le projet est à rendre avant le 26 janvier avant 23h59. Les étudiants devront indiquer leurs noms dans le document final.

Le projet est à envoyer par mail avec la mention [projetINDEX] au début de l'objet du mail.

Le mail est à envoyer à l'adresse suivante : [youness.el-marhraoui02@univ-paris8.fr](mailto:youness.el-marhraoui02@univ-paris8.fr)