

Projet BDDR

Objectif:

A partir d'un jeu de données, réaliser une application web permettant de passer des requêtes paramétrées et de visualiser les réponses.

Contexte:

Pendant la pandémie du COVID-19, de nombreux articles de recherche ont été publiés pour mieux comprendre l'impact de la maladie aux différentes échelles. L'urgence de la situation sanitaire nous a aussi mené à une urgence académique: celle de la classification et l'organisation de la recherche (et résultats) générés.

C'est dans ce contexte là que le jeu de données CORD-19 (The COVID-19 Open Research Dataset) est né. Il est conçu pour faciliter le développement de systèmes d'exploration de textes et de recherche d'informations grâce à sa riche collection de métadonnées et de documents structurés en texte intégral.

Sujet:

L'application à développer est destinée aux chercheurs intéressés de trouver rapidement des publications pertinentes dans leur domaine de recherche. Elle doit leur proposer un jeu de formulaires leur permettant d'explorer et de visualiser les données et, si possible, proposer des liens vers des sites web qui pourraient chercher l'article même.

Suite aux données, c'est une arborescence des fichiers CSVs regroupés par thématiques.

Socle minimal des requêtes:

1. Liste d'articles par thématique et sous-thématique.
2. Histogramme d'articles publiés par date, semaine, mois.
3. Liste de thématiques
4. Nombre de publications par labo/institution.
5. Liste de journaux par nombre et type de publications.

Toute requête supplémentaire est bienvenue.

Besoins techniques:

Votre application doit être entièrement développée avec le framework web **Django** et le SGBD **PostgreSQL**. Pour le peuplement de la base de données l'utilisation de Python est conseillée mais vous pouvez choisir la technologie à utiliser. La gestion des versions avec **Git** (sur Github) est obligatoire, ainsi que des mises à jour fréquentes (1 par semaine, au minimum).

Modalité du projet:

- Projet à réaliser en binôme.
- Jeux de données sur [Kaggle](#)
(Attention, une fois téléchargé, supprimez le repertoire "cord_embeddings") >
- On considère comme livrables de votre projet:
 - a. Le code développé pour l'application
 - b. Le code du(des) script de peuplement de la base de données,
 - c. Le script de création de la base de données
 - d. Le schéma du modèle de la base de données (en format DBM ou PDF)..
 - e. Le rapport final.

- f. Un fichier README pour indiquer comment lancer votre projet dans une autre machine.
- Pour le rapport:
 - a. 5 pages max.
 - b. Trois parties bien définies: Contexte, base de données et application.

Calendrier:

Semaine	Activité(s)
4	<ul style="list-style-type: none">● Renseigner la composition de son binôme et l'URL du repo Git● donner les droits de consultation du repo git aux correcteurs● mise en place de l'environnement de développement
4 ~ 7	<ul style="list-style-type: none">● étude des données, nettoyage● conception du modèle de données
8	<ul style="list-style-type: none">● point d'étape : chaque binôme expose son modèle de données (le schéma et le script doivent être à jour)
9 ~ 12	<ul style="list-style-type: none">● peuplement de la base de données
13	<ul style="list-style-type: none">● évaluation des scripts de peuplement à partir du dépôt git
14 ~ 17	<ul style="list-style-type: none">● développement de l'application web● rédaction du petit rapport, synthétisant les idées suivies,● les idées non suivies, les difficultés, les surprises, et le travail réalisé.
18	<ul style="list-style-type: none">● vérifier que le dépôt Git est complet● évaluation du projet - soutenance