

# Projet de fin de semestre

## Description générale

Ce projet consiste à donner des outils d'analyse sur un data set portant sur certaines *issues* du site web github.

## Provenance des données et méthodes d'obtention

Le site web github.com fourni une api rest sur l'url <https://api.github.com/search/>

l'exemple de données a été téléchargé avec la commande

```
wget "https://api.github.com/search/issues?q={python,sklearn}&page,per_page,sort,order}"
```

Une documentation de l'api est accessible sur l'url : <https://developer.github.com/v3/>

## Structures attendues

- Charger de l'ensemble des données en une base de données mongoDB.
- Créer d'une base de données neo4j avec comme nœuds :
  - le dépôt git (*repository\_url*)
  - l'auteur de l'issue.
  - l'issue
  - le *author association*

## Livrables attendus

Gérer l'ordre des posts d'une même issue avec de liens directionnel dans la base neo4j.

Faire une fonction python qui affiche l'ensemble d'un post.

Faire une fonction python qui affiche les utilisateurs liés à un poste.

Faire une fonction python qui affiche les participations d'un utilisateur ordonnés par date.

N'hésitez pas à nous montrer toutes autres informations ou fonctions utiles sur ces données.

## Documents attendu :

- Le code complet transférer sur le site github, celui-ci doit marcher en démarrant un docker-compose et en lançant un minimum de commande, voir aucune commande.
- Un Readme clair spécifiant :
  - La structure des bases de données.
  - L'utilité de chaque script et leur fonctionnement.
  - Les commandes à lancer pour exécuter les différentes fonctions pythons.
- Une présentation détaillant :
  - le set de données.

- les bases de données et leur structure.
- les réponses aux livrables.
- d'autres choses intéressantes s'il y en a.