

TD 2 : Manipulation de fichiers en Python

1 Extraction d'informations

Cet exercice vise à manipuler un fichier .csv qui est un fichier de type tableur dont les informations sont séparées par des virgules. Le but est d'extraire les informations de ce fichier et de les récupérer dans un dictionnaire. À cause du format du fichier, vous l'ouvrirez avec l'encodage utf-8-sig.

Le fichier Hospitals.csv est un fichier qui répertorie les hôpitaux de Détroit, chaque ligne correspond à un hôpital et ses informations (nom, nombre de lits, adresse etc.) séparées par des virgules.

- Ouvrir le fichier et lire le fichier **ligne par ligne**.
- Parcourir chaque ligne et séparer au niveau des virgules avec la fonction `split()`.
- Créer un dictionnaire vide pour chaque ligne et placer chaque élément de la liste obtenue dans le dictionnaire avec les clés suivantes : nom, lits, adresse, contact, code. Ci-dessous, le dictionnaire contenant les informations de la première ligne.

```
1 {'nom': 'St. John Hospital & Medical Ctr. - Detroit', 'lits': '572', 'adresse': '22101 Moross Rd.', 'contact': 'St. John Health System', 'code': '48236'}
```

- Ajouter chaque dictionnaire dans une liste. Vous devez obtenir une liste comme ci-dessous dont les éléments sont les dictionnaires :

```
1 [{'nom': 'Department of Veterans Affairs Detroit Medical Center', 'lits': '432', 'adresse': '4646 John R Rd.', 'contact': 'Department of Veteran Affair', 'code': '48201'},
2 {'nom': 'St. John Hospital & Medical Ctr. - Detroit', 'lits': '572', 'adresse': '22101 Moross Rd.', 'contact': 'St. John Health System', 'code': '48236'},
3 ...]
```

- Afficher le nombre total d'hôpitaux répertoriés
- Afficher le nombre total de lit
- Afficher le nombre d'hôpitaux ayant pour code 48207
- Afficher le nombre de lits des hôpitaux ayant pour code 48201

2 Comparaison de textes de langues différentes

Vous avez à votre disposition deux extraits des actes du Parlement Européen, l'un en français `europarl_fr.txt`, l'autre en anglais `europarl_en.txt`. Dans ce projet, vous allez comparer les caractéristiques lexicales des deux langues, entre autres :

- La taille du texte en caractères/en mots
- La longueur des mots en caractères
- La longueur des phrases en mots/en caractères
- Les mots les plus fréquents : sont-ils exactement les mêmes ? Si non, pourquoi ?

Appliquez des pré-traitements adaptés à votre analyse et illustrez-la avec un ou plusieurs graphiques.

Utilisez les bibliothèques `matplotlib` ou `seaborn` pour créer vos graphiques.