

SAÉ105 : Traiter des données (partie 2)

Dictionnaire (le type `dict`)

Un dictionnaire (type `dict`) est une collection de `paires` d'éléments ordonnés :

- Chaque paire associe une `clé` à une `valeur`.
- Un dictionnaire est identifié par une référence.
- La déclaration d'un dictionnaire se fait par l'instruction : `dicto = {cle1 : valeur1, cle2 : valeur2, ..., cleN : valeurN}`.
- Le nombre de paires (éléments) d'un dictionnaire `dicto` est donné par la fonction `len(dicto)`.
- On accède en lecture à la valeur qui est associée à une clé du dictionnaire `dicto` par l'expression `valeur = dicto[cle]`.
- On modifie la valeur qui est associée à une clé du dictionnaire `dicto` par l'expression `dicto[cle] = valeur`.
- On ajoute une nouvelle paire au dictionnaire `dicto` par l'expression `dicto[cle] = valeur`, si `cle` n'est pas déjà une clé du dictionnaire `dicto`.

Problème

On suppose le même environnement de travail que vous avez utilisé pour la première partie de la SAÉ. On considère aussi la même définition pour les mots et les chaîne de caractères.

1. Ecrire une fonction python `mots_dans_fichiers_rep(rep)` qui recense tous les fichiers contenus dans le répertoire `rep` dans une liste `liste_fichiers_rep`. Cette fonction doit aussi recenser tous les mots utilisés dans l'ensemble des fichiers de la liste `liste_fichiers_rep` et compter le nombre d'utilisations de chacun de ces mots dans chacun des fichiers. La fonction prendra en entrée le nom du répertoire `rep` (on suppose que le répertoire existe et contient au moins un fichier texte) et renverra une liste dont chaque élément est une liste de deux éléments : le premier élément est le nom du fichier, le deuxième élément est un dictionnaire dont les clés sont les mots utilisés et les valeurs sont nombre d'utilisation de ces mots dans le fichier correspondant.
2. Ecrire une fonction python `apparition_mots_rep(rep)` qui calcule la fréquence d'apparition des 15 mots apparaissant le plus fréquemment dans chacun des fichiers contenus dans le répertoire `rep`. La fonction prendra en entrée le nom du répertoire `rep` (on suppose que le répertoire existe et contient au moins un fichier texte) et renverra une liste dont chaque élément est un dictionnaire de 15 paires : les clés du dictionnaire sont les 15 mots apparaissant le plus fréquemment dans le fichier correspondant et la valeur associée à chaque clé est la fréquence d'apparition du mot dans le fichier. La liste renvoyée par cette dernière fonction doit être enregistrée dans un fichier texte.

3. Ecrire une fonction python `compte_mot_rep(mot, rep)` qui calcule la fréquence d'apparition d'un mot donné en entrée par l'utilisateur dans chacun des fichiers contenus dans le répertoire `rep`. La fonction prendra en entrée le mot donné par l'utilisateur `mot` ainsi que le nom du répertoire `rep` (on suppose que le répertoire existe et contient au moins un fichier texte) et renverra un dictionnaire dont les clés sont les noms des fichiers contenus dans `rep` et les valeurs sont les fréquences d'apparition du mot donné par l'utilisateur le fichier. Afficher l'ensemble de ces fréquences dans un graphique (penser à utiliser la fonction `matplotlib.pyplot.bar()`)

On considère les deux fichiers `texte_1.txt` et `text_2.txt` qui contiennent des tweets (récupérés sur Twitter aléatoirement entre le 27/11 et 08/12). Analyser la forme de ces données.

4. Ecrire une fonction python `lecture_tweet(rep)` qui permet de lire tous les tweets contenus dans tous les fichiers contenus dans le répertoire `rep`. La fonction prendra en entrée le nom du répertoire `rep` (on suppose que le répertoire existe et contient au moins un fichier texte) et renverra une liste dont chaque élément est une liste de deux éléments : le premier élément est la date de publication du tweet et le deuxième élément le texte du tweet.
5. Modifier le programme précédant pour créer une fonction `compte_tweet(rep)` qui compte le nombre de tweet par journée. La fonction prendra en entrée le nom du répertoire `rep` (on suppose que le répertoire existe et contient au moins un fichier texte) et renverra un graphique représentant le nombre de tweet par jour entre le 27/11 et 08/12.
6. Ecrire une fonction python `compte_mot_tweet_rep(mot, rep)` qui calcule la fréquence d'apparition d'un mot donné en entrée par l'utilisateur par jour sur l'ensemble des tweets contenus dans les fichiers contenus dans le répertoire `rep`. La fonction prendra en entrée le mot donné par l'utilisateur `mot` ainsi que le nom du répertoire `rep` (on suppose que le répertoire existe et contient au moins un fichier texte) et renverra un dictionnaire dont les clés sont les dates des tweets et les valeurs sont les fréquences d'apparition du mot donné par l'utilisateur à la date correspondante. Afficher l'ensemble de ces fréquences dans un graphique.