

TP 06

Tableaux, chaînes de caractères et fichiers*Savoirs et compétences :*

- ☐ AA.C9 : Choisir un type de données en fonction d'un problème à résoudre
- ☐ AA.S11 : Manipulation de quelques structures de données.
- ☐ AA.S12 : Fichiers

Consignes

Attention : suivez précisément ces instructions. Vous enverrez à votre enseignant un fichier d'extension `.py` (script Python) nommé

`tp06_durif_kleim.py`,

où les noms de vos enseignants sont à remplacer par ceux des membres du binôme. Le nom de ce fichier ne devra comporter ni espace, ni accent, ni apostrophe, ni majuscule. Dans ce fichier, vous respecterez les consignes suivantes.

- Écrivez d'abord en commentaires (ligne débutant par #), le titre du TP, les noms et prénoms des étudiants du groupe.
- Commencez chaque question par son numéro écrit en commentaires.
- Les questions demandant une réponse écrite seront rédigées en commentaires.
- Les questions demandant une réponse sous forme de fonction ou de script respecteront pointilleusement les noms de variables et de fonctions demandés.

Les figures produites aux différentes questions concernées devront respecter les noms demandés avec la présence des noms des membres du binôme. Le nom de ces fichiers ne devra comporter ni espace, ni accent, ni apostrophe, ni majuscule.

Activité 1 : Analyse d'un dipôle électrique

Copier et coller le fichier "dipole_electrique.txt" que vous trouverez sur le site de la classe dans votre répertoire personnel. Celui-ci contient des données numériques, obtenues en TP d'électronique, correspondant à différentes mesures de l'intensité traversant un dipôle et de la tension entre ses bornes. On souhaite récupérer ces données et les exploiter.

Ouvrir le fichier "dipole_electrique.txt" à l'aide d'un éditeur de texte (de type bloc-notes) et observer son contenu.

Q 1 : Écrire un programme Python que vous appellerez `lit_dipole` et qui prendra en argument une variable de type chaîne de caractère qui sera le nom du fichier de données. Cette fonction renverra deux listes notées `I` et `U`, l'une contenant les intensités mesurées, et l'autre contenant les tensions mesurées.

Q 2 : Créer une fonction que l'on notera `tracer_dipole` qui prendra en argument deux listes représentant respectivement l'intensité `I` et `U`. Cette fonction permettra de représenter graphiquement les couples de points `(I,U)`. On sauvegardera la figure avec le nom 'tp06_noms_q02.png'

Q 3 : Conjecturer alors une relation reliant la tension aux bornes du dipôle à l'intensité le traversant.

Activité 2 : Lecture et analyses des caractères d'un texte

Ouvrir le fichier `complot_contre_lamerique.txt` dans python (on prendra soin de nommer la variable contenant cet objet). Ce fichier contient un extrait du livre de Philip Roth, *Complot contre l'Amérique*.

Q 4 : Que fait chacune des méthodes `read()`, `readline()` et `readlines()` ? Quels sont les types des valeurs que chacune de ces fonctions renvoient ?

Q 5 : Que représentent les symboles `\t` et `\n` ?

Q 6 : Écrire une fonction Python `carac(nom_de_fichier)` qui renvoie un tableau contenant le nombre de caractères de chaque ligne du fichier `nom_de_fichier`, en excluant les symboles `\t` et `\n`.

Q 7 : Écrire une fonction Python `somme_carac(nom_de_fichier)` qui renvoie le nombre de caractères total contenu dans le fichier `nom_de_fichier`, en excluant les symboles `\t` et `\n`.

Indication : attention au type de `nom_de_fichier` !

Q 8 : Écrire une fonction Python `compte_carac(carac, nom_de_fichier)` qui renvoie pour un caractère de l'alphabet (noté `carac`) son nombre d'occurrences contenu dans le fichier `nom_de_fichier` sans tenir compte de la casse.

Q 9 : Écrire une fonction Python `stat_carac(nom_de_fichier)` qui renvoie une liste de 26 éléments donnant le nombre d'occurrences de chaque lettre de l'alphabet contenu dans le texte `nom_de_fichier` sans tenir compte de la casse.

Indication :

On pourra créer une variable globale contenant tous les caractères de l'alphabet :

`alphabet='abcdefghijklmnopqrstuvwxyzt'.`

Q 10 : Écrire une fonction Python `tracer_occurrences(carac, nom_de_fichier)` qui trace en fonction du numéro de la lettre dans l'alphabet son nombre d'occurrences dans le fichier `nom_de_fichier`. La figure obtenue avec l'extrait du livre *complot contre l'Amérique* de Philip Roth sera sauvegardée sous le nom `tp06_vosnoms_q10.png`.

Q 11 : Écrire une fonction Python `tracer_stat_occurrences(nom_de_fichier)` qui trace en fonction de chaque lettre dans l'alphabet sa fréquence d'apparition dans le fichier `nom_de_fichier` en pourcentage par rapport au nombre total de caractères présents en excluant les symboles `\t` et `\n`. Le tracé devra avoir l'apparence d'un histogramme. La figure obtenue avec l'extrait du livre *complot contre l'Amérique* de Philip Roth sera sauvegardée sous le nom `tp06_vosnoms_q11.png`.

Indication :

- Pour faire apparaître les lettres en abscisses, on pourra utiliser la fonction `xticks` du module `matplotlib.pyplot`.
- Pour tracer les barres l'histogramme on pourra utiliser tout simplement des segments verticaux.

On donne le fichier `frequence_wikipedia.csv` téléchargeable sur le site de la classe qui contient les fréquences en pourcentage d'apparition des 26 lettres de l'alphabet dans tous les articles français de Wikipedia. On pourra l'ouvrir avec un éditeur de texte pour observer son contenu.

Q 12 : Écrire une fonction Python `stat_carac_wikipedia(nom_de_fichier)` qui à partir du fichier `nom_de_fichier` contenant les statistiques de fréquences de caractères (sous le même format que dans le fichier `frequence_wikipedia.csv`) renvoie une liste donnant en fonction de la position de la lettre dans l'alphabet sa fréquence en %.

Q 13 : Écrire une fonction Python `tracer_stat_wikipedia()` qui trace en fonction de chaque lettre dans l'alphabet, sa fréquence d'apparition dans tous les articles français de wikipedia en pourcentage par rapport au nombre total de caractères présents. Le tracé devra avoir l'apparence d'un histogramme. La figure obtenue sera sauvegardée sous le nom `tp06_vosnoms_q13.png`.

Q 14 : Utiliser les différentes fonctions réalisées précédemment pour comparer les fréquences de caractères entre l'extrait de *complot contre l'Amérique* de Philip Roth et celles issues de l'ensemble des articles français de wikipedia. Conclure.