

Python pour les Humanités

Consignes Examen

Jean Barré - Janvier 2024

Exercice 1 : Manipulations de base 6 points

Librairies utilisables (exclusif) : random

1. Travaux préliminaires :

- (a) Sans utiliser `max()`, créer une fonction `maxval` qui prend en entrée deux entiers et retourne le plus grand des deux.
- (b) Sans utiliser `min()`, créer une fonction `minval` qui prend en entrée deux entiers et retourne le plus petit des deux.
- (c) Générer une liste `ListAlea` composée de 20 nombres aléatoires compris entre 0 et 30 (en cas de non réussite, définir une liste de 20 entiers compris entre 0 et 30, non triée et avec répétitions).
- (d) Créer une fonction `Existence` qui prend en entrée deux listes et retourne le dictionnaire des fréquences d'apparition des entiers de la liste `{clef = nombre : valeur = nombre d'apparition}`.
- (e) Enregistrer ces fonctions et cette liste dans un fichier `maths_exam.py` et l'importer dans votre notebook.

Soit la liste suivante :

```
ListExam1=[1,"chameau",0.9,16,7,2,0,12,32,"elephant",(0,0),3,25,5,4,18,27,22,1,32]
```

- 2. Trouver le maximum et le minimum de cette liste en utilisant les fonctions précédentes. Toutes les exceptions devront être traitées dans le fichier `maths_exam.py`.
- 3. En utilisant les fonctions précédentes, pour chaque élément de `ListExam1`, dire s'il se trouve ou non dans `ListAlea`.

Exercice 2 : Matrice et crible d'Ératosthène 6 points

1-A Échiquier

- Créer une matrice échiquier (des 1 et des 0 alternés) de taille 8×8 , de deux façons différentes
 1. En vous servant de slices
 2. En vous servant de la fonction `tile`

1-B Piège lors d'une extraction 2D

- Définir la matrice $M = \begin{pmatrix} 1 & 5 & 9 & 13 & 17 \\ 2 & 6 & 10 & 14 & 18 \\ 3 & 7 & 11 & 15 & 19 \\ 4 & 8 & 12 & 16 & 20 \end{pmatrix}$
- En **extraire** la matrice $\begin{pmatrix} 6 & 18 & 10 \\ 7 & 19 & 11 \\ 5 & 17 & 9 \end{pmatrix}$

1-C Crible d'Erathosthène

On souhaite implémenter un crible d'Ératosthène pour trouver les nombres premiers inférieurs à $N = 1000$.

- Partir d'un array de booléens de taille $N + 1$, tous égaux à `True`.
- Mettre 0 et 1 à `False` car ils ne sont pas premiers.
- Pour chaque entier k entre 2 et \sqrt{N} :
 - Si k est premier, on passe ses multiples (entre k^2 et N) à `False`.
- On retourne la liste des entiers premiers.

Exercice 3 : Balzac et Sand

8 points

Restriction : Les fonctions lambda ne sont pas autorisées

L'objectif de cet exercice est d'évaluer la différence des styles entre Honoré de Balzac et George Sand.

Version courte : Créer un module capable de calculer la fréquence relative des N=100 mots les plus fréquents, puis de calculer la distance euclidienne entre chaque texte.

La ligne de commande suivante doit réaliser le travail :

```
python examen_ex3.py -i chemin_dossier_txt -a output_frequencies.csv  
-b output_distances.csv -n 100 -sw mots_outils.json
```

Version longue :

- Créer une fonction qui ouvre les fichiers txt du dossier 'data'.
- Une autre qui calcule la fréquence relative des N=100 mots les plus fréquents (relative à la longueur du roman), retourne et sauve le résultat sous la forme d'un dataframe (10 lignes = 10 romans, 100 colonnes = 100 mots).
- Faire la même chose, mais prendre en compte le fichier `mots_outils_fr.json` qui est une liste des mots outils pour le français. Cette fonction compte la fréquence des mots si et seulement s'ils ne sont pas dans la liste de mots-outils.
- Écrire une fonction qui calcule la distance euclidienne entre chaque texte, retourne et sauve le résultat sous la forme d'un dataframe.
- Interpréter les résultats avec et sans mots-outils.
- Implémenter un module Python `examen_ex3.py` qui utilise les fonctions précédemment définies et sauvegarde les fichiers résultants. La ligne de commande suivante doit s'exécuter correctement :

```
python examen_ex3.py -i chemin_dossier_txt -a output_frequencies.csv  
-b output_distances.csv -n 100 -sw mots_outils.json
```
- Bonus : Visualiser les résultats des distances sous la forme de plusieurs graphes (une carte de chaleur des distance et une autre), expliquer votre choix de visualisation.