

Dernière mise à jour	Informatique	Claire GAUDY - Denis DEFAUCHY
10/11/2022	3 - Modules - Bibliothèques	TD 3-1 – Lecture dans un fichier

Informatique

3

Modules - Bibliothèques

TD 3-1

Lecture dans un fichier

Dernière mise à jour	Informatique	Claire GAUDY - Denis DEFAUCHY
10/11/2022	3 - Modules - Bibliothèques	TD 3-1 – Lecture dans un fichier

Exercice 1: Nombres premiers

Vous avez à votre disposition le fichier « 3-1 - TD - Lecture dans un fichier - Premiers 1000.txt » dont [voici le lien](#). Il contient les 1000 premiers nombres premiers.

Question 1: Calculer la somme des 1000 premiers nombres premiers

Question 2: Déterminer le 684ème nombre premier

Exercice 2: Joueurs de Tennis

Vous avez à votre disposition le fichier « 3-1 - TD - Lecture dans un fichier - Tennis homme.txt » dont [voici le lien](#). Il contient des informations sur des joueurs de tennis avec leur classement en 2013.

Question 1: Déterminer le nombre de joueurs français dans ce classement.

On cherche à calculer toutes ces moyennes :

- Moyenne des scores de tous les joueurs (≈ 964)
- Moyenne des scores des 100 premiers joueurs (≈ 1519)
- Moyenne des scores des joueurs français (≈ 1109)
- Moyenne des rangs des joueurs dont le prénom se termine par un r (≈ 110)

Question 2: Mettre en place le code permettant de les déterminer

Exercice 3: Admissibles

Vous avez à votre disposition le fichier « 3-1 - TD - Lecture dans un fichier - Admissibles.txt » dont [voici le lien](#). Il contient la liste des élèves de MP qui se sont présentés à un concours. Il contient exclusivement des lignes de la forme :

- 1234 TOURNESOL Tryphon Admissible 2
- 43210 HADDOCK Archibald Éliminé -

la 1ère valeur est le numéro du candidat et la dernière valeur est le numéro de série d'oral quand l'élève est admissible. Il y a 4 séries d'oral. Le caractère de séparation des 4 champs est la tabulation '\t'.

Question 1: Construire une liste $[n, n_1, n_2, n_3, n_4]$ où n est le nombre total de candidats et n_i est le nombre de candidats pour la série d'oral i . Combien y a-t-il de candidats admissibles ?

Remarque : On parcourra une seule fois le fichier

Dernière mise à jour	Informatique	Claire GAUDY - Denis DEFAUCHY
10/11/2022	3 - Modules - Bibliothèques	TD 3-1 – Lecture dans un fichier

Exercice 4: Comptage par dictionnaire - Hamlet

Vous avez à votre disposition le fichier « 3-1 - TD - Lecture dans un fichier - Hamlet français.txt » dont [voici le lien](#). Il contient des informations sur des joueurs de tennis avec leur classement en 2013.

Question 1: En utilisant un dictionnaire, dénombrer les interventions des différents personnages lorsqu'ils parlent seuls

Remarques :

- Exemple à la ligne 754 où ils sont 2 à intervenir : « HORATIO et MARCELLUS »
- Il semble qu'il y ait quelques erreurs dans le fichier texte, ligne 392 on trouve 2 auteurs : « HAMLET. - Elle me donne un bon conseil. (Sort le seigneur.) HORATIO. - Vous perdrez ce pari, monseigneur. ». On ne tiendra pas compte du second auteur

Pour la suite, on ne s'intéresse qu'aux cas où les personnages parlent seuls. On cherche à répondre à ces questions :

- Combien de fois parle HAMLET ?
- Et LA_REINE ?
- Combien y a-t-il de personnages qui parlent ?
- Combien y a-t-il de prises de parole en tout ?

Question 2: Ecrire les instructions Python permettant de répondre aux questions posées

Dernière mise à jour	Informatique	Claire GAUDY - Denis DEFAUCHY
10/11/2022	3 - Modules - Bibliothèques	TD 3-1 – Lecture dans un fichier

Exercice 5: Comptage par dictionnaire – ADN

Question 1: Ecrire une fonction `Fichier_to_str(nom_fichier)` qui prend en argument le nom du chemin d'accès d'un fichier, et renvoie une seule chaîne de caractère contenant tous les caractères du fichier hormis ceux de la 1^{ère} ligne.

Nous avons à notre disposition des fichiers d'extension « .fna » qui sont principalement utilisés pour les fichiers FASTA DNA And Protein Sequence Alignment. Par chance, ils s'ouvrent comme des fichiers texte sous Python.

Vous les avez à disposition :

- « 1-3 - TD - Lecture dans un fichier - NC_000913.fna » dont [voici le lien](#), contient le génome de la bactérie *Escherichia coli*
- « 1-3 - TD - Lecture dans un fichier - NC_001133.fna » dont [voici le lien](#), contient la séquence du chromosome I de la levure du boulanger *Saccharomyces cerevisiae*

Question 2: Créer la fonction `Compte_mots_2_lettres(texte)` qui, étant donnée une chaîne de caractère texte, renvoie un dictionnaire de tous les mots de 2 lettres avec leurs nombres d'apparitions

Question 3: Déterminer les mots de 2 lettres ainsi que leurs occurrences dans cette séquence

Vérifier que le mot « GC » apparaît 8919 fois

Question 4: Créer la fonction `Compte_mots_k_lettres(texte,k)` qui, étant donnée une chaîne de caractère texte, renvoie un dictionnaire de tous les mots de k lettres avec leurs nombres d'apparitions

Question 5: Combien de fois apparaît la protéine « CCC » dans le génome de la bactérie *Escherichia coli*

Réponse : 47775

Question 6: Ecrire une fonction `Maxi_dico(dico)` qui, étant donné un dictionnaire dont les valeurs sont des nombres, renvoie la/les clé(s) dont la valeur est maximale

Vérifier :

```
>>> Maxi_dico({'t1' : 32, 't2' : 2, 't3' : 50, 't4' : 23, 't5' : 50})
['t3', 't5']
```

Question 7: Déterminer le mot de 3 lettres qui apparaît le plus souvent dans le génome de la bactérie *Escherichia coli*

Réponse : « CGC » 115695 fois