

TP de Langages de script n° 7 : manipulation de fichiers textuels et expressions régulières

I) Avant de commencer

Votre source principale

Si vous ne connaissez pas encore ce site, naviguez vers <http://docs.python.org/py3k/> pour regarder toute la documentation officielle de PYTHON. En particulier, nous conseillons fortement de parcourir l'excellent *Tutorial* pour vos révisions et pour l'approfondissement. Et quand vous cherchez les renseignements sur un module et ses fonctions, rendez-vous dans *Library Reference*

II) Et maintenant les exercices

Exercice 1 : La disparition

Écrire un script qui, prenant en entrée un nom de fichier, en crée une “copie” dans laquelle toutes les occurrences du caractère **e** ont été supprimées.

Exercice 2 : Cent mille milliards de poèmes de Raymond Queneau

Le fichier `vers-queneau.txt` contient une séquence de 14 groupements de 10 vers séparés par des lignes vides.

1. Écrire un script PYTHON chargeant le fichier `vers-queneau.txt` dans un dictionnaire associant aux entiers de 0 à 13, la liste des vers de chaque groupement.
2. Modifier le script pour générer un poème de 14 vers en prenant un vers aléatoirement dans chaque groupement. Utiliser pour cela le module `random` de la bibliothèque standard de PYTHON (<http://docs.python.org/py3k/library/random.html>). Ne pas oublier de signer Raymond Queneau à la fin du poème.

Exercice 3 : Expression régulières : prise en main

Pour cet exercice, on va utiliser le module `re` (<http://docs.python.org/py3k/library/re.html>, voir aussi <http://docs.python.org/py3k/howto/regex.html>).

Vous pouvez tester vos expressions sur des petites chaînes d'abord, mais aussi sur le texte de Flaubert qui se trouve dans le fichier `educ.txt`.

1. Générez les listes de toutes les occurrences des mots qui contiennent la lettre **k** ou **w**, toutes les années de 19-ème siècle, toutes les formes du verbe **aimer**. Modifiez vos expressions pour ne chercher que les derniers mots de la ligne.
2. Si ce n'est pas encore fait pour l'exercice 1 du TP 5, écrivez en utilisant le module `re` une fonction `decoupe(phrase)` qui prenne en argument la chaîne `phrase` et renvoie une liste des mots de cette chaîne (ne contenant pas le mot vide). Les mots peuvent être délimités par des espaces ou par tout signe de ponctuation (virgule, point, etc.).
3. Affichez les numéros de lignes qui contiennent les années de 19-ème siècle.

4. (optionnel) Comment trouver tous les mots qui contiennent 2 fois les mêmes trois lettres consécutives (comme **barbarie** ou **entrent**). Les phrases qui contiennent 2 fois le même mot.
5. Affichez les numéros de lignes et les positions dans ces lignes de toutes les formes du verbe **aimer**.

Exercice 4 : Modifications de l'Éducation Sentimentale

On continue de travailler sur le fichier `educ.txt`.

1. Ce texte est formé de chapitres annoncés par une ligne contenant **Chapitre N** où N est le numéro du chapitre. Écrire un script chargeant le fichier en mémoire et créant un fichier pour chaque chapitre, intitulé `chapitre-i.txt`.
2. Modifier le script pour qu'il remplace le prénom "Frédéric" par "Gustave".
3. Modifier le script pour qu'il remplace tous les mots commençant par "aim" suivi d'un suffixe par le mot "détest" suivi du même suffixe.
4. Modifier le script pour qu'il extraie les dialogues de chaque chapitre dans un fichier nommé `chapitre-dialogues-i.txt`.