T.P. n° 12

Les fréquences des lettres utilisées dans les mots d'une langue varient beaucoup. La lettre 'e' est ainsi la plus fréquente en français (autour de 15 %), alors que la lettre 'w' a une fréquence bien plus faible (autour de 0,5%).
Ce TP consiste à écrire un programme qui compte les fréquences des lettres dans un très long texte, ce qui nous permettra d'avoir une bonne fiabilité statistique. Par exemple, vous pourrez récupérer « *le Rouge et le Noir* », de Stendhal, sur Moodle ou ici : http://www.gutenberg.org/ebooks/798.txt.utf-8 en supprimant dans ce dernier cas, avec un éditeur de texte, le début et la fin du document qui sont en anglais.

a) Écrire un programme qui compte la fréquence de la lettre 'E' (ou 'e') dans ce texte. Pour cela, il vous faut parcourir chacune des lignes du fichier et pour chacune d'elles, parcourir chacun des caractères. Vous devriez trouver environ 15%.

b) Pour calculer la fréquence de chacune des lettres de 'A' à 'Z', on va utiliser un tableau à 26 cases, une pour chacune des lettres (0 pour 'A', 1 pour 'B', etc.). Chaque case sera initialisée à 0. Il vous faudra ensuite parcourir chacun des caractères comme précédemment et, lorsqu'il s'agit d'une lettre, incrémenter la valeur de la case correspondante du tableau . Pour prendre en compte les lettres minuscules et majuscules, vous pouvez utiliser la fonction `Character.toUpperCase(<caractère>)` qui renvoie le caractère en majuscule.
A la fin, il faut parcourir le tableau, calculer les fréquences et les afficher. Voici ce que vous devriez trouver :
A: environ 8%
B: environ 1%
C: environ 3%
D: environ 4%
E: environ 15%
...
Quelle est la seconde lettre la plus fréquente en français ?

c) Écrire ces résultats dans le fichier `frequence.txt` plutôt qu'à l'écran.

d) Quel prénom apparaît le plus fréquemment dans le roman : Julien ou Mathilde ?

e) **Bonus difficile** : afficher tous les mots commençant par une majuscule.