

TP05

L'objectif de ce TP est de vous faire pratiquer l'utilisation des dictionnaires

1) Fréquence des mots et distance de document

Programme qui va dénombrer les mots d'un texte et les stocker dans un dictionnaire. Chaque mot sera une clé du dictionnaire et la fréquence à laquelle il apparaît dans le texte sera sa valeur.

Ainsi dans l'exemple d'un texte composé de "To be or not to be" on aurait un dictionnaire (en mettant tout en minuscule) :

```
{"to": 2, "be": 2, "or": 1, "not": 1}
```

Réaliser ce travail sur les 4 textes fournis.

Une fois les fréquences de mots établies pour chaque texte, demandez à l'utilisateur d'entrer un mot. Il faudra ensuite calculer la distance/proximité du document (à quel point le mot recherché est proche du texte) pour chaque document.

Pour calculer la distance :

```
round(Nombre_de_fois_le_mot_recherché / Nombre_de_mot_du_texte, 2)
```

Astuces

Vous n'allez traiter que les lettres "autorisées" fournies dans le code initial. Vu que vous allez traiter le fichier caractère par caractère, vous allez recomposer les mots puis les ajouter dans votre dictionnaire. Si vous tombez sur un espace " ", une apostrophe "'" ou un retour à la ligne "\n", cela signifie que votre mot est terminé et que vous pouvez l'ajouter dans votre dictionnaire.

Toutes les autres ponctuations ne sont pas traitées. Mettez toutes les lettres en minuscules lors du traitement, cela évitera d'avoir des mots différents juste pour une majuscule.

La variable chemin contient le chemin d'accès au répertoire où se trouvent les fichiers textes.

La variable files contient une liste ayant chaque nom de fichier comme élément. Cela vous permet de boucler dessus pour ouvrir un fichier après l'autre.

Exemple de sortie :

```
Entrez le mot que vous recherchez : plages
Le texte 'texte1.txt' comporte 12 fois les mots ['plages'] sur 19 mots.
Ce qui fait une proximité de 0.63
Le texte 'texte2.txt' comporte 0 fois les mots ['plages'] sur 5516 mots.
Ce qui fait une proximité de 0.0
Le texte 'texte3.txt' comporte 0 fois les mots ['plages'] sur 8684 mots.
Ce qui fait une proximité de 0.0
Le texte 'texte4.txt' comporte 0 fois les mots ['plages'] sur 22334 mots.
Ce qui fait une proximité de 0.0
Entrez le mot que vous recherchez : science
Le texte 'texte1.txt' comporte 0 fois les mots ['science'] sur 19 mots.
Ce qui fait une proximité de 0.0
Le texte 'texte2.txt' comporte 2 fois les mots ['science'] sur 5516 mots.
Ce qui fait une proximité de 0.0
Le texte 'texte3.txt' comporte 1 fois les mots ['science'] sur 8684 mots.
Ce qui fait une proximité de 0.0
Le texte 'texte4.txt' comporte 234 fois les mots ['science'] sur 22334 mots.
Ce qui fait une proximité de 0.01
Entrez le mot que vous recherchez : internet
Le texte 'texte1.txt' comporte 0 fois les mots ['internet'] sur 19 mots.
Ce qui fait une proximité de 0.0
Le texte 'texte2.txt' comporte 89 fois les mots ['internet'] sur 5516 mots.
Ce qui fait une proximité de 0.02
Le texte 'texte3.txt' comporte 0 fois les mots ['internet'] sur 8684 mots.
Ce qui fait une proximité de 0.0
Le texte 'texte4.txt' comporte 0 fois les mots ['internet'] sur 22334 mots.
Ce qui fait une proximité de 0.0
Entrez le mot que vous recherchez : panda
Le texte 'texte1.txt' comporte 0 fois les mots ['panda'] sur 19 mots.
Ce qui fait une proximité de 0.0
Le texte 'texte2.txt' comporte 0 fois les mots ['panda'] sur 5516 mots.
Ce qui fait une proximité de 0.0
Le texte 'texte3.txt' comporte 0 fois les mots ['panda'] sur 8684 mots.
Ce qui fait une proximité de 0.0
Le texte 'texte4.txt' comporte 0 fois les mots ['panda'] sur 22334 mots.
Ce qui fait une proximité de 0.0
```