

Fréquence d'une lettre dans un texte.



On va préparer ici la partie statistique du déchiffrement d'un texte crypté.

A partir d'un texte donné, on va relever les fréquences d'apparition des lettres dans ce texte.

On utilisera deux types de présentation des résultats : sous forme de liste et sous forme de dictionnaire.

Dans un script FREQLET

On considère le texte suivant : `chaine="stage ti python"` (ou tout autre texte, cela peut-être un texte beaucoup plus long téléchargé).

1°) Ecrire une fonction `espace` qui prend comme argument une chaîne de caractères `ch` et qui renvoie le nombre d'espace dans cette chaîne.

```
PYTHON SHELL
>>> from FREQLET import *
>>> espace(chaine)
2
```

2°) Ecrire une fonction `flettre` qui prend comme arguments une chaîne de caractères `chaine` et un caractère `c` et qui renvoie la fréquence d'apparition du caractère `c` dans le texte `chaine`.

```
PYTHON SHELL
>>> flettre(chaine,"a")
0.07692307692307693
>>> flettre(chaine,"t")
0.2307692307692308
>>> flettre(chaine,"z")
0.0
```

3°) Ecrire une fonction `freq` qui prend comme argument une chaîne de caractères `chaine` et qui renvoie la liste des fréquences des lettres "a", "b", ..., "z" dans `chaine`.

```
PYTHON SHELL
>>> freq(chaine)
[0.07692307692307693, 0.0, 0.0, 0.0, 0.07692307692307693, 0.0, 0.07692307692307693, 0.07692307692307693, 0.07692307692307693, 0.07692307692307693, 0.0, 0.0, 0.07692307692307693, 0.07692307692307693, 0.07692307692307693, 0.07692307692307693, 0.0, 0.0, 0.07692307692307693, 0.2307692307692308, 0.0, 0.0, 0.0, 0.0, 0.07692307692307693, 0.0]
```

4°) Ecrire une fonction `pourcent` qui prend comme argument une chaîne de caractères `chaine` et qui renvoie le dictionnaire des pourcentages d'apparition des lettres "a", "b", ..., "z" dans `chaine` (on écrira seulement les lettres dont la fréquence associée est non nulle).

```
PYTHON SHELL
>>> pourcent(chaine)
{'e': 7.69, 'i': 7.69, 'g': 7.69, 'h': 7.69, 'a': 7.69, 't': 23.08, 'o': 7.69, 'y': 7.69, 'p': 7.69, 's': 7.69, 'n': 7.69}
```



Fréquence d'une lettre dans un texte.



Fonction espace

1°) On va initialiser un compteur `s` à 0.

Puis on va parcourir toutes les lettres de `chaîne` à l'aide d'une boucle `for`.

A chaque fois qu'une lettre est un espace alors on incrémente le compteur `s` de 1.

```
ÉDITEUR : FREQLET
LIGNE DU SCRIPT 0009
def espace(ch):
    s=0
    for c in ch:
        if c==" ":
            s=s+1
    return s
```

Fonction flettre

2°) On peut envisager 2 façons d'écrire cette fonction :

- En utilisant les fonctions et méthodes de Python

`len` permet d'obtenir la taille d'une liste ou le nombre de caractères d'une chaîne (il faudra lui retirer le nombre d'espaces dans le texte).

`chaîne.count(c)` renvoie le nombre de fois où la chaîne `c` est présente dans `chaîne`.

Cette façon de faire est très rapide.

- On peut aussi utiliser un compteur `s` et, dans une boucle `for`, parcourir tous les caractères de `chaîne` et ajouter 1 au compteur dès que le caractère rencontré est égal à `c`.

Cette méthode est plus longue mais elle utilise des notions de programmation classiques.

```
ÉDITEUR : FREQLET
LIGNE DU SCRIPT 0019
def flettre(chaîne,c):
    n=len(chaîne)-espace(chaîne)
    s=chaîne.count(c)
    return s/n
```

```
ÉDITEUR : FREQLET
LIGNE DU SCRIPT 0024
def flettre(chaîne,c):
    s=0
    n=len(chaîne)-espace(chaîne)
    for i in chaîne:
        if i==c:
            s=s+1
    return s/n
```



Fréquence d'une lettre dans un texte.



Fonction freq

3°) `liste` contiendra la liste des fréquences des lettres a, b, c, ..., z dans notre chaîne.

On va donc initialiser `liste` comme une liste vide. Cette fois on n'utilisera pas `liste=[]` mais `liste=list()` qui sont deux instructions équivalentes.

`alpha` est la chaîne de caractère contenant toutes les lettres de l'alphabet.

`for i,c in enumerate(alpha)` signifie que `c` va parcourir tous les caractères de la chaîne `alpha` et `i` sera l'indice associé.

La première valeur de `c` sera "a" et `i` vaudra 0.

La seconde valeur de `c` sera "b" et `i` vaudra 1, etc...

A l'aide de la fonction `flettre` on calcule à chaque tour de boucle `f` la fréquence d'apparition de la lettre `c` (qui vaut "a" au début puis "b",...) dans le texte `chaîne`.

Et enfin on ajoute cette fréquence dans la liste des fréquences qu'on a appelé `liste`.

```
ÉDITEUR : FREQLET
LIGNE DU SCRIPT 0031
def freq(chaîne):
    liste=list()
    alpha="abcdefghijklmnopqrstu
xyz"
    for i,c in enumerate(alpha):
        f=flettre(chaîne,c)
        liste.append(f)
    return liste
```

Notion de dictionnaire en Python.

Un dictionnaire est un objet de type `dict`.

Un élément du dictionnaire est caractérisé par sa **clef** et sa **valeur** :

La **clef** aussi bien que sa **valeur** associée peut-être une chaîne de caractères ou un nombre.

Cela ressemble un peu aux listes, sauf que la clef peut-être une chaîne de caractères ou même un `float` !

```
PYTHON SHELL
>>> dico=dict()
>>> dico["pomme"]=17
>>> dico["poire"]="comice"
>>> dico[8]="huit"
>>> dico[10]=20
>>> dico
{'poire': 'comice', 8: 'huit', 10: 20, 'pomme': 17}
```

Fonction pourcent

4°) Cette fonction est sensiblement identique à `freq`.

On a initialisé notre dictionnaire `dico` en écrivant `dico=dict()`

On aurait pu aussi écrire `dico={ }` mais c'était moins parlant et on pouvait confondre avec le type `set` (ensemble).

On a seulement besoin de parcourir tous les caractères de l'alphabet, calculer la fréquence et la multiplier par 100 pour obtenir un pourcentage.

On a utilisé l'instruction `round` pour obtenir 2 décimales de précision.

A la fin on a testé si `f` était non nul afin de l'ajouter dans le dictionnaire `dico` avec une **clef** correspondant à la lettre de l'alphabet et comme **valeur** le pourcentage associé.

```
ÉDITEUR : FREQLET
LIGNE DU SCRIPT 0039
def pourcent(chaîne):
    dico=dict()
    alpha="abcdefghijklmnopqrstu
xyz"
    for c in alpha:
        f=flettre(chaîne,c)
        f=round(100*f,2)
        if f!=0:
            dico[c]=f
    return dico
```

