

DÉCRYPTAGE STATISTIQUE

Vous êtes agent secret et vous venez de recevoir ce mail crypté (ou codé). Vous savez que c'est un poème écrit en français :

```
YR PNEER QR Y ULCBGRAHFR
RFG RTNY FV WR AR Z NOHFR
N YN FBZZR QRF PNEERF
QRF QRHK NHGERF PBGRF
```

Votre mission si vous l'acceptez, consiste à déchiffrer (ou décoder, ou décrypter) ce message. Dans votre panoplie d'agent secret, vous disposez de Python, mais son utilisation n'est pas indispensable!

Pour commencer, on peut recopier le message en Python (le code "+" sert à rajouter des morceaux de texte, au texte courant).

```
code='YR PNEER QR Y ULCBGRAHFR '
code+='RFG RTNY FV WR AR Z NOHFR '
code+='N YN FBZZR QRF PNEERF '
code+='QRF QRHK NHGERF PBGRF '
```

Pour faire des statistiques sur le message codé, on crée un objet *message* composé des lettres du message codé (dans un tableau) puis un ensemble *lettres* pour éviter d'écrire plusieurs fois la même lettre. Cet objet est l'ensemble des lettres du message.

```
message=[x for x in code]
lettres=set(message)
```

Pour décoder le message, on va utiliser l'algorithme d'AL KINDI (801-873), consistant à comparer la fréquence d'apparition des lettres codées avec la fréquence des lettres dans la langue française. En effet, si une lettre apparaît souvent dans la langue française, on s'attend à ce qu'elle apparaisse souvent dans un message codé.

Pour faire le tableau d'effectifs, on peut demander à Python de compter les lettres du message :

```
for y in lettres:
    print(y+'-->'+str(message.count(y)))
```

Remplir le tableau d'effectifs suivant (on ne compte pas les espaces puisque ce ne sont pas des lettres). ***On n'a pas besoin de Python pour ce TP, il suffit de compter les lettres !*** :

lettres	A	B	C	E	F	G	H	K	L	N	O	P	Q	R	U	T	V	W	Y	Z
effectifs																				

On compare avec les fréquences des lettres dans la langue française, dont voici les plus fréquentes :

lettres	E	S	A	I	T	N	R	U	L	O	D	C	P
fréquences (%)	14,7	7,9	7,6	7,5	7,2	7,1	6,6	6,3	5,5	5,4	3,7	3,3	3

Quelle lettre code le "E" ?

De même, quelle lettre code le "S" ?

Et quelle lettre code le "A" ?

Pour les lettres suivantes, les fréquences sont trop proches pour être aussi certain. Pour continuer à décoder, on a donc intérêt à utiliser des propriétés linguistiques comme le fait qu'un mot long commence souvent par une consonne, qu'il y a peu de mots de deux lettres surtout au début de la phrase etc.

Décoder alors le message :



Pour occuper les plus rapides

Tester la variante suivante du script, qui utilise l'objet "dictionnaire" de Python :

```
effectifs={}  
for x in code:  
    effectifs[x]=code.count(x)  
  
print(effectifs)
```

Si le temps le permet, effectuer une recherche par Internet sur le livre "la disparition" de Georges Perec. Quelle est la particularité statistique de ce livre ?