Tp ALGORITHME des k plus proches voisins



L'objectif de ce travail est d'implémenter un algorithme qui permette à partir d'une liste échantillon de mots français et allemands, de pouvoir prédire si un mot mystère quelconque est plus de consonnance française ou allemande.

1. Travail préparatoire sur papier

Pour réaliser cette prédiction, on décide ici de tenir compte uniquement de la longueur du mot mystère et du nombre de voyelles qui le composent. On souhaite prédire l'origine des mots suivants : « anoure », « ohrwurm » et « vitrine ».

1- Compléter les tableaux suivants :

Echantillon:

Mots	Longueur	Nombre de voyelles	Origine
hiver			français
loup			français
reines			français
nordique			français
winter			allemand
schloss			allemand
koniginnen			allemand
nordisch			allemand

Mots mystères dont on souhaite déterminer l'origine :

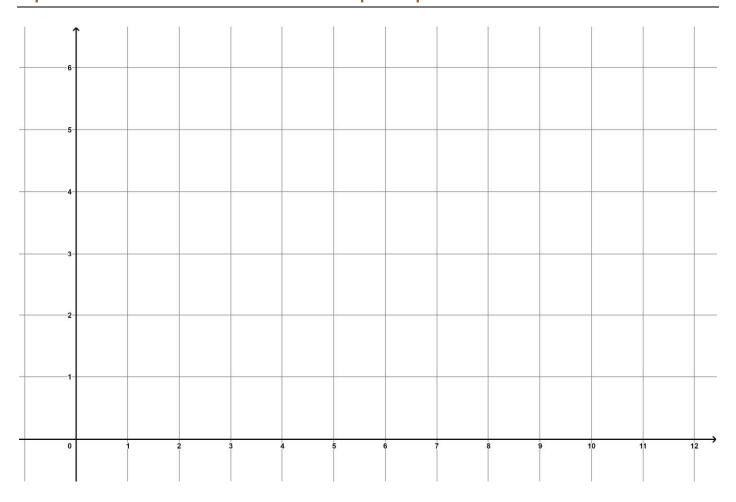
Mots	Longueur	Nombre de voyelles
anoure		
ohrwurm		
vitrine		

2- Représenter les données de l'échantillon et des mots mystères sur le graphique qui suit et qui représente le nombre de voyelles des mots en fonction de la longueur : utiliser des points bleus pour les mots français, rouge pour les allemands et verts pour ceux dont on souhaite déterminer l'origine.

ou:

ALGORITHME des k plus proches voisins





3- En utilisant des arguments graphiques, donner si possible, une prédiction quant à l'origine des mots « anoure », « ohrwurm » et « vitrine » :

Mots	Origine
anoure	
ohrwurm	
vitrine	

2. Implémentation d'un algorithme de prédiction de la langue d'origine d'un mot

Le code donné sur la page suivante permet à l'exécution d'obtenir :

```
>>> (executing file "langue_voisins.py")
nombre de voisins : 3
fr fr al

>>> (executing file "langue_voisins.py")
nombre de voisins : 5
fr fr al fr fr
```

On peut en déduire que le mot « *anoure* » est d'origine française. Ce résultat est en accord avec l'étude sur papier réalisée dans la partie 1.