

1 TD 11 : Arbre et trie

(correction page ??)

Abordé lors de cette séance	
programmation	recherche textuelle
algorithme	structure de données, trie

On veut vérifier qu'un mot appartient à une liste de mots de trois façons différentes.

Première demi-heure : Construction d'une liste aléatoire

Plutôt que de charger un texte en mémoire, on construit des mots aléatoirement.

Exercice 1 :

A partir de la fonction suivante, construit une liste aléatoire de 10000 mots d'une longueur de 20 lettres.

```
import random
def mot_alea (l) :
    l = [ chr(97+random.randint(0,25)) for i in range(l) ]
    return "".join(l)
```

Par la suite, on construit une seconde liste plus courte de mots à chercher qu'on ajoute à la première.

Exercice 2 :

Les listes ont une méthode `index` qui permet de retrouver la position d'un mot (elle effectue donc une recherche). Le module `time` possède une fonction `clock`. Utiliser cette fonction pour mesurer le temps du code suivant :

```
for k in cherche :
    i = mots.index(k)
```

Seconde demi-heure : Recherche dichotomique

Lors des premières séances de TD, on a implémenté la recherche dichotomique.

Exercice 3 :

Mesurer le temps pris par la fonction.

Quels sont les coûts des deux recherches en fonction du nombre de mots dans la liste ?

Troisième demi-heure : Trie

Le coût des deux recherches dépendent du nombre de mots dans la liste. Il est possible de construire une structure de données qui fait dépendre le coût de cette recherche du nombre de lettres différentes (26) et de la longueur maximale des mots.

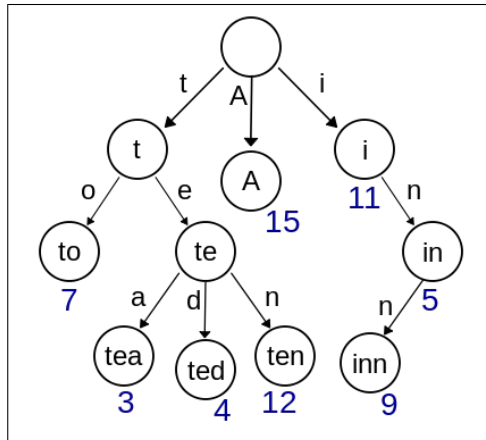


FIGURE 1 : Illustration d'un trie, figure extraite de Wikipédia (<http://en.wikipedia.org/wiki/Trie>).

Exercice 4 :

Construire cette structure.

Quatrième demi-heure : Recherche dans un trie

Exercice 5 :

Une fois qu'on a la structure, il faut aussi écrire la fonction qui recherche un mot dans le trie.

Exercice 6 :

Il ne reste plus qu'à mesurer le temps.

Pour aller plus loin ou pour ceux qui ont fini plus tôt

Modifier le programme pour trace la courbe des différents coûts en fonction de la taille de la liste de mots et de la longueur des mots.

Remarques