# Recherche d’un mot dans un texte

La recherche d’un mot dans un texte, c’est-à-dire d’une *petite* chaîne de caractères dans une *grande* chaîne de caractères est un problème classique qui figure au programme d’Informatique Pour Tous.

## 1) Poser le problème, le résoudre en trichant un peu.

On dispose d’un texte T qui est une chaîne de caractères et d’un mot M qui est également une chaîne de caractères et on veut savoir si M est dans T, autrement dit : si M est une sous-chaîne de T.

Python dispose d’une fonction qui permet de résoudre simplement le problème : in.

Question 1 : proposer une fonction Python utilisant in et qui résolve le problème posé.

**Limites de cette approche :**

- on n’a aucune idée de la complexité de la fonction in. Deux pistes : lire sa documentation et la mesurer en testant la fonction.
- Python est un langage de haut niveau, ce qui signifie qu’il dispose d’outils sophistiqués que n’ont pas tous les langages, par exemple : in. (Ces outils ont un coût).

## 2) Résolution naïve du problème

Question 2 : proposer une fonction auxiliaire qui teste si M commence au caractère i de T.

Question 3 : estimer le coût de la fonction précédente.

Question 4 : en utlisant la fonction auxiliaire, proposer une fonction qui résolve le problème posé.

Question 5 : estimer le coût de la fonction précédente.

**Limites de cette approche :**

le coût est très élevé. Les meilleurs algorithmes ont des complexité en O(len(T)). Les implémentations dans les programmes usuels combinent plusieurs de ces algorithmes.

Question 6 : pourquoi n’implémente-t-on pas uniquement un algorithme dont la complexité est en O(len(T)) ?

### 3) Algorithme de Boyer-Moore

Illustrons le principe de l'algorithme de Boyer-Moore sur un exemple : on cherche à tester l'appartenance du mot « carambar » dans un texte T.

- étape 1 : a-t-on la 8è lettre de T qui est r ?
- étape 2 : si oui, on teste la 7è lettre et ainsi de suite pour voir si « carambar » correspond aux 8 premières lettres de T.
- étape 3 : sinon, la 8è lettre de T est-elle une autre lettre de « carambar »? Si oui, on teste pour voir si « carambar » commence à l'une des lettres parmi la 2è et la 8è.
- étape 4 : sinon, on recommence mais à partir de la 9è lettre de T.

Question 7 : Programmer l'algorithme de Boyer-Moore.

### 4) Verdict ?

Question 8 : Faire des tests de comparaison de l'efficacité des différentes solutions.

On pourra utiliser le code suivant :

```
import string
import random
def chaineAlea(n):
    # n est un int
    # renvoie une chaîne de caractères aléatoire de taille n
    str = string.ascii_lowercase
    return ''.join(random.choice(str) for i in range(n))
```