

# Recherche d'un motif dans un texte

## Problématique générale

On dispose d'un texte assimilé à une seule chaîne de caractères et d'un motif, qui est lui aussi une chaîne de caractères supposée de longueur inférieure à celle du texte.

Il s'agit de déterminer les positions éventuelles du motif dans le texte.

On assimilera les chaînes de caractères à des tableaux indicés à partir de 0.

## Exemple élémentaire

Prenons le texte `on dit bonjour .` et le motif `on` alors la réponse est la liste de deux positions `[0,8]`.

## Le cœur du problème.

### Problème 1

On cherche maintenant un algorithme permettant de résoudre la problématique générale exposée au début de ce document.

*Un fil pour vous accompagner dans votre recherche.*

1. Traduisez sous forme algorithmique la recherche qui a conduit à trouver le résultat pour l'exemple élémentaire.
2. Testez cet algorithme avec les exemples suivants :
  - (a) texte : `abracadabra` et `blabla` et motif : `abr`
  - (b) texte : `abababa` et motif : `aba`
3. Cet algorithme va-t-il toujours s'arrêter ?

Quel est un ordre de grandeur du nombre d'opérations à réaliser pour compléter le tableau des positions éventuelles du motif dans le texte ?
4. Essayez de trouver une amélioration de cet algorithme au sens du nombre d'instructions nécessaires.
5. Testez votre proposition avec les trois exemples proposés ou d'autres.

## Le problème complémentaire

On considère un mot, c'est-à-dire une chaîne de caractères. On appelle préfixes du mot toutes les chaînes de caractères croissantes construites à partir du premier caractère du mot. On définit aussi les suffixes du mot.

## Exemple

Le mot `mot` a ainsi pour préfixes `m`, `mo` et `mot` et pour suffixes `t`, `ot` et `mot`.

On étend ces notions aux chaînes de caractères.

### Problème 2

On demande de déterminer un algorithme qui, pour une chaîne de caractères désignée par `motif`, produit le tableau contenant, pour tout nombre entier naturel  $k$  compris entre 1 et la longueur de la chaîne `motif`, la plus grande valeur de l'entier naturel  $j$  tel que

$$\text{motif}[0 : j] = \text{motif}[k - j : k]$$

Autrement dit, il s'agit, pour chaque préfixe de la chaîne de caractères, de trouver la longueur du plus grand de ses préfixes égal à un de ses suffixes.

### **Exemple**

Avec la chaîne de caractères baobab qui a pour préfixes b, ba, bao, baob, baoba et baobab on trouve le tableau [0, 0, 0, 1, 2, 1].

*Un fil pour vous accompagner dans votre recherche.*

1. Essayez de comprendre la construction du tableau de cet exemple.
2. Écrivez tous les préfixes de la chaîne de caractères abracadabra puis, face à chacun de ces préfixes, notez la réponse à la demande.  
Déduisez alors le tableau associé à cette chaîne de caractères.
3. Recommencez avec la chaîne de caractères bébé puis avec bonbons et enfin avec bonhomme.
4. Élaborez alors un algorithme qui réponde à la demande.
5. Testez cet algorithme avec chacun des quatre exemples précédents.
6. Cet algorithme va-t-il toujours s'arrêter ?  
Quel est un ordre de grandeur du nombre d'opérations à réaliser pour compléter le tableau ?