

Numérique et Sciences Informatiques  
Chapitre XIII - Recherche textuelle  
Travaux Dirigés 22

## I. Recherche naïve

La première idée de recherche d'un motif dans un texte est la suivante :

- Je compare le 1er caractère du motif avec le 1er caractère du texte.
- Si les deux caractères sont égaux, je compare le deuxième caractère du motif avec le deuxième caractère du texte. Et ainsi de suite.
- Sinon je décale le motif d'un caractère et je réitère le procédé en commençant avec le 2ème caractère du texte.

Implémenter cette algorithme dans une fonction `recherche(motif, texte)` de paramètres deux chaînes de caractères et qui renvoie l'indice (dans le texte) à partir duquel on retrouve le motif dans le texte.

## II. Algorithme de Boyer-Moore

L'algorithme de Boyer-Moore se fait en trois étapes :

- prétraitement du motif.
- comparaison du motif et de la chaîne en commençant par la droite du motif.
- trouver le bon décalage du motif pour la comparaison suivante.

1. Exemple Nous allons appliquer l'algorithme de Boyer-Moore sur la chaîne de caractères `ATAGACACAATAT` avec le motif `ATATAC`.

- (a) Le prétraitement du motif  $M$  consiste en la création d'un tableau indiquant la dernière position de chaque symbole du motif dans le sous-motif (la sous-chaîne de caractères)  $M[0..i]$  avec  $0 \leq i < \text{longueur}(M)$ . Compléter le tableau ci-dessous.

i	A	T	C	G
0				
1				
2				
3				
4				
5				

- (b) On compare le motif et la chaîne en partant de la droite du motif.  
`ATAGACACAATATAC`  
`ATATAC`

Sur quel caractère la comparaison renvoie `False` ? (position dans le texte et position dans le motif)

- (c) On trouve le bon décalage en regardant le prétraitement du motif.

De combien faut-il décaler le motif vers la droite, en étant sûr qu'on n'oublie pas des comparaisons ?

- (d) Refaire les question b et c jusqu'à trouver le motif dans le texte s'il est présent.  
Y a-t-il un lien entre le décalage et le tableau de prétraitement ?

2. Implémentation Nous avons vu qu'il y avait un lien entre le décalage et le tableau de prétraitement, donc il est plus judicieux d'indiquer les décalages dans le tableau de prétraitement.

- (a) Implémenter une fonction `pretraitement(motif)` où `motif` est une chaîne de caractères, qui renvoie le tableau des décalages des caractères qui composent le motif.
- (b) Implémenter une fonction `recherche_BOYER_Moore(motif, texte)` de paramètres `motif` et `texte`, deux chaînes de caractères, et qui renvoie la position du motif dans le texte s'il est présent, `None` sinon. *Remarque : il pourra y avoir d'autres paramètres à la fonction.*