

Exercices - Aléa 2019

Exercice I : Transformée inverse (Burrows-Wheeler)

Soit le texte $t = \text{urattg\$uuooo}$.

1. Calculer la fonction LF (last-to-first).
2. Calculer la transformée inverse de t .

Exercice II : Recherche et localisation avec la transformée de Burrows-Wheeler

3. Calculer la transformée de Burrows-Wheeler du texte aabaabaabba .
4. En s'inspirant de l'algorithme pour la transformée inverse, donner un algorithme qui permette de trouver le nombre d'occurrence pour un motif u donné. Dérouler (sur papier) votre algorithme pour les motifs suivants : ab, bba, aaa, aba .
5. De quoi a-t-on besoin en plus pour localiser la position des occurrences dans le texte ?
6. Donner le code pour compter le nombre d'occurrences d'un motif utilisant la fonction LF .

Exercice III : Localisation par les bords ou les préfixes

On rappelle qu'un mot u est un bord d'un mot x s'il est différent de x et est à la fois préfixe et suffixe de x . Si x n'est pas le mot vide, on note $Bord(x)$ le plus long bord de x .

Soit x un mot non vide de taille m , la table des bords de x est un tableau d'entiers TB , de taille m tel que pour $i \in \{0, \dots, m\}$:

$$TB[0] = -1 \quad \text{et, pour } 1 \leq i \leq m, TB[i] = |Bord(x_0 \cdots x_{m-1})|.$$

7. Pour deux mots x et y , décrire de quelle manière on peut utiliser directement la table des bords pour le mot $x\$y$ (où le symbole $\$$ n'apparaît ni dans x ni dans y) afin de localiser toutes les occurrences du mot x dans le texte y .
8. Quelle est complexité en temps de cette méthode ? Pensez-vous qu'il y ait des situations où elle est intéressante ?

Exercice IV : Indexation avec table de suffixes

9. Soit le texte TGCTGTTTGCTG .
 - (a) Construire sa table des suffixes TS et sa table HTR .
 - (b) Déterminer également, l'inverse de sa table des suffixes. Donner le code correspondant et la complexité du calcul de cette table.
10. Calculer à l'aide de la table des suffixes le(s) plus long(s) facteur(s) répété(s). Quelle est la complexité de votre méthode ?
11. Déterminer les répétitions supermaximales du texte TGCTGTTTGCTG .

Remarque : une répétition supermaximale est un facteur répété qui n'est facteur propre d'aucune autre répétition.