

Algorithmique - Licence M/I

TD 1

20 septembre 2004

On considère un texte t qui est un mot de longueur n (grande) et un motif w qui est mot de longueur m généralement petite devant n . Le i -ième caractère d'un mot u est noté u_i ou $u[i]$ et la longueur de u est notée $|u|$. Ainsi, $u = u_0u_1\dots u_{|u|-1}$.

Exercice 1: Un motif w apparaît comme facteur d'un texte t s'il existe $i \in [0; n - m]$ tel que, pour tout $j \in [0; m - 1]$, $w_j = t_{i+j}$. On dira alors que le motif apparaît à la position i . Donner un algorithme naïf qui permet de tester si un motif est facteur d'un texte.

Exercice 2: Quel est le nombre maximal de comparaisons effectuées par cet algorithme. Donner un exemple de texte et de motif atteignant ce nombre.

Exercice 3: Comment améliorer cet algorithme si tous les caractères du motif sont différents ?

Afin d'améliorer l'algorithme naïf, si on arrive dans une situation où $w_j \neq t_{i+j}$, on veut pouvoir tenir compte des informations dont on dispose. (Par exemple, on peut savoir que, pour tout $k < j$, $w_k \neq t_{i+k}$. Dans ce but on définit le *bord* d'un mot comme le plus long préfixe propre du mot qui est aussi un suffixe : v est un bord de u si et seulement si

- i) $|v| < |u|$,
- ii) pour tout $i < |v|$, $v_i = u_i = u_{|u|-|v|+i}$,
- iii) $|v|$ maximal.

Exercice 4: Donner les bords de *ababa*, *baba* et *babbbabb*.

Exercice 5: Montrer que les bords vérifient la récurrence suivante :

$$\text{bord}(ua) = \begin{cases} \text{bord}(u)a & \text{si } \text{bord}(u)a \text{ préfixe de } u \\ \text{bord}(\text{bord}(u)a) & \text{sinon.} \end{cases}$$

En déduire un algorithme qui permette de calculer pour tout mot u le tableau B tel que, pour tout $i < |u|$, $B[i]$ est la longueur du bord de $u_0\dots u_i$.

Exercice 6: Utiliser le tableau B pour améliorer l'algorithme de l'exercice 1.

Un *sous-mot* (en anglais *subsequence*) d'un mot u est un mot s tel qu'il existe une suite strictement croissante $(i_j)_{j < |s|}$ de $[0; |u| - 1]$ tel que, pour tout $j < |s|$, $s_j = u_{i_j}$. Par exemple *crabe* est un sous-mot de *carabine*. On note $\text{plsm}(u,v)$ le plus long sous-mot commun à u et v .

Exercice 7: Exprimer $\text{plsm}(ua,vb)$ en fonction de $\text{plsm}(u,vb)$, $\text{plsm}(ua,v)$ et $\text{plsm}(u,v)$.

Exercice 8: Calculer le plsm de *bdcaba* et *abcabdab* en utilisant les règles précédentes.

Exercice 9: Donner un algorithme pour calculer $\text{plsm}(u,v)$.