

# TP : Algorithme de Knuth-Morris-Pratt

Dans ce TP, on considérera des mots de type `string`.

Soit  $u = u_0 \cdots u_{n-1}$  un mot. On appelle *bord* de  $u$  tout préfixe différent de  $u$  qui soit aussi suffixe de  $u$ . On note  $\beta(u)$  le bord de longueur maximale.

1. Écrire une fonction `bord_bof : string -> string` qui calcule le bord maximal d'un mot.
2. Soit  $k$  le plus petit entier tel que  $\beta^k(u) = \varepsilon$ . Montrer que pour toute lettre  $a$ ,  $\beta(ua)$  est le plus long préfixe de  $x$  dans  $\{\varepsilon\} \cup \{\beta(u)a, \dots, \beta^k(u)a\}$ .
3. En déduire que pour tout mot  $u$  et toute lettre  $a$ ,

$$\beta(xa) = \begin{cases} \beta(x)a & \text{si } \beta(xa) \text{ est un préfixe de } x \\ \beta(\beta(x)a) & \text{sinon} \end{cases} .$$

La partie principale de l'algorithme consiste à calculer la longueur des bords maximaux de tous les préfixes d'un mot. On notera alors, pour un mot  $u$ ,  $\rho_u(k)$  la longueur du bord maximal du mot  $u_0 \cdots u_{k-1}$ . On posera  $\rho_u(0) = -1$ .

4. Soit  $k < n$ . On considère les entiers  $j_1 = \rho_u(k)$ ,  $j_2 = \rho_u(j_1)$ ,  $\dots$ ,  $j_{k+1} = b(j_k)$ .  
Montrer que  $j_{k+1} = -1$ . Soit donc  $\alpha$  le plus petit entier tel que  $u_{j_\alpha} = u_k$  ou  $j_{\alpha} = -1$ . Montrer que  $b(k+1) = j_\alpha + 1$ .
5. En déduire une fonction qui calcule le tableau des longueurs des bords des préfixes d'un mot.

L'algorithme de Knuth-Morris-Pratt fonctionne de la manière suivante pour chercher le mot  $u$  dans le texte  $s$  :

- On construit le tableau  $b$  des longueurs des bords maximaux des préfixes du mot recherché.
  - On compare les lettres du mot et du texte.
  - Si la comparaison échoue aux lettres  $u_i \neq s_{j+i}$  se présentent deux cas :
    - si  $b(i+1) > 0$  ou  $i = 0$ , alors on décale le texte de  $i+1$  positions, et la comparaison du mot reprend à 0
    - sinon,  $u_i$  n'est pas une lettre d'un facteur, on reprend l'analyse du texte à l'indice  $j+i-b(i)$  et celle du mot à  $b(i)$ .
6. Écrire la fonction qui applique l'algorithme de Knuth-Morris-Pratt.
  7. Deux mots  $u$  et  $v$  sont dits conjugués s'il existe  $s$  et  $t$  tels que  $u = ts$  et  $v = st$ .  
Montrer que deux mots sont conjugués si et seulement s'ils ont même longueur, et  $u$  facteur de  $vv$ .
  8. En déduire une fonction qui vérifie que deux mots sont conjugués.