TD d'algorithmique numéro 13 - Recherche de motifs, suite
Daniel Hirschkoff et Natacha Portier
semaine du 10 au 15 janvier 2000

## Question 1 — ⋆ Knuth Morris Pratt ⋆

Cet algorithme permet d'éviter le calcul de la fonction de transition de l'automate fini. On le remplace par le calcul d'une nouvelle fonction, la $f$onction préfixe, qui exprime la correspondance entre le motif et ses propres décalages.

### Question 1.1 — ⋆ définition de la fonction préfixe ⋆

La $f$onction préfixe $\pi$ pour le motif $P$ de longueur $m$ est une fonction de $\{1; 2 \ldots m\}$ dans $\{0; 1 \ldots m-1\}$ définie par :

$$\pi[q] = \max\{k \mid k < q \text{ et } P_k \sqsupset P_q\}$$

Si les $q$ premières lettres du motif ont déjà été trouvées dans le texte, c'est-à-dire s'il existe un décalage $s$ tel que $M[s+1] \ldots M[s+q] = P[1] \ldots P[q] = P_q$, alors le décalage $s$ est potentiellement valide (il reste à vérifier que la fin du motif est éventuellement présente) et le prochain décalage potentiellement valide est $s' = s + (q - \pi[q])$.

Calculer la fonction $\pi$ pour le motif *ababaca*.

### Question 1.2 — ⋆ calcul de la fonction préfixe ⋆

Montrons la validité de la procédure :

**Calcul_Fonction_Préfixe**$(P)$ :=
  $m :=$ *longueur*$[P]$
  $\pi[1] := 0$
  $k := 0$
  `pour` $q$ `de` 2 `à` $m$
    `faire tant que` $k > 0 \ \wedge\ P[k+1] \neq P[q]$
      `faire` $k := \pi[k]$
    `si` $P[k+1] = P[q]$ `alors` $k := k+1$
    $\pi[q] := k$
Retourner $\pi$

On pose $\pi^*[q] = \{q, \pi[q], \pi^2[q], \ldots, \pi^t[q]\}$, la séquence s'arrêtant lorsque l'on atteint $\pi^t[q] = 0$.

1. Montrer que pour $q = 1, 2, \ldots, m$ :

$$\pi^*[q] \;=\; \{k : P_k \sqsupset P_q\}\,.$$

2. Montrer que pour $q = 1, 2, \ldots, m$, si $\pi[q] > 0$ alors $\pi[q] - 1 \in \pi^*[q-1]$.

   Pour $q = 2, 3, \ldots, m$ on définit le sous ensemble $E_{q-1} \subseteq \pi^*[q-1]$ par

$$E_{q-1} = \{k \mid k \in \pi^*[q-1] \text{ et } P[k+1] = P[q]\}$$

3. Montrer que pour $q = 2, 3, \ldots, m$ :

$$\pi[q] = \begin{cases} 0 & \text{si } E_{q-1} = \emptyset \\ 1 + \max\{k \in E_{q-1}\} & \text{si } E_{q-1} \neq \emptyset \end{cases}$$

4. En déduire la correction de **Calcul_Fonction_Préfixe**.

**Question 1.3** ⋆ complexité de la fonction préfixe ⋆

Utiliser une analyse amortie pour donner le temps d'exécution de la fonction **Calcul_Fonction_Préfixe**. On prendra comme potentiel la valeur de l'entier $k$.

**Question 1.4** ⋆ procédure KMP ⋆

**KMP**(T,P) :=
  $n :=$ *longueur*$[T]$
  $m :=$ *longueur*$[P]$
  $\pi :=$ **Calcul_Fonction_Préfixe**$(P)$
  $q := 0$
  `pour` $i$ `de` 1 `à` $n$
    `faire tant que` $q > 0 \ \wedge \ P[q+1] \neq T[i]$
      `faire` $q := \pi[q]$
    `si` $P[q+1] = T[i]$ `alors` $q := q + 1$
    `si` $q = m$ `alors` **Imprimer** *Le motif apparaît avec le décalage* $i - m$
                $q := \pi[q]$

Utiliser un argument similaire à celui de la question précédente pour calculer le temps d'exécution du reste de la fonction **KMP**.

En déduire le temps d'exécution de l'algorithme.

**Question 1.5** ⋆ Validité de l'algorithme de Knuth Morris Pratt ⋆

La procédure **KMP** simule la procédure de recherche par automate fini. On donne à $q$ la valeur $\delta(q, T[i])$ non pas en utilisant une valeur stockée mais en la recalculant si nécessaire.

Montrer que soit $\delta(q, T[i]) = 0$ soit $\delta(q, T[i]) - 1 \in \pi^*[q]$.

En déduire la correction de l'algorithme **KMP**.