

Option Informatique en Spé MP et MP*

Nombre maximal de sous-mots d'un mot : le corrigé

Question 1 • Ce ne sont pas les mots qui manquent : double, décor, ovale...

Question 2 • Réflexivité: l'identité I est une injection croissante de $\llbracket 1, m \rrbracket$ dans lui-même telle que $x_{I(i)} = x_i$ pour tout $i \in \llbracket 1, m \rrbracket$.

• Antisymétrie: soient x et y vérifiant $x \sqsubseteq y$ et $y \sqsubseteq x$. Soient s une injection croissante de $\llbracket 1, m \rrbracket$ dans $\llbracket 1, n \rrbracket$ telle que $y_{s(i)} = x_i$ pour tout $i \in \llbracket 1, m \rrbracket$, et t une injection croissante de $\llbracket 1, n \rrbracket$ dans $\llbracket 1, m \rrbracket$ telle que $x_{t(j)} = y_j$ pour tout $j \in \llbracket 1, n \rrbracket$. On a nécessairement $n = m$; puis s , injection croissante de $\llbracket 1, m \rrbracket$ dans lui-même est l'application identique, si bien que $y = x$.

• Transitivité: soient x, y et z vérifiant $x \sqsubseteq y$ et $y \sqsubseteq z$. Notons $p = |z|$, s une injection croissante de $\llbracket 1, m \rrbracket$ dans $\llbracket 1, n \rrbracket$ telle que $y_{s(i)} = x_i$ pour tout $i \in \llbracket 1, m \rrbracket$, et t une injection croissante de $\llbracket 1, n \rrbracket$ dans $\llbracket 1, p \rrbracket$ telle que $z_{t(j)} = y_j$ pour tout $j \in \llbracket 1, n \rrbracket$. Alors $z_{(t \circ s)(i)} = x_i$ pour tout $i \in \llbracket 1, m \rrbracket$, et $t \circ s$ est une injection croissante. Ainsi $x \sqsubseteq z$.

• Compatibilité avec la concaténation: supposons $x \sqsubseteq y$, et soit s une injection croissante de $\llbracket 1, m \rrbracket$ dans $\llbracket 1, n \rrbracket$ adaptée. Soit z de longueur p ; alors $t : \llbracket 1, m+p \rrbracket \mapsto \llbracket 1, n+p \rrbracket$ définie par $t(i) = i$ pour $i \in \llbracket 1, p \rrbracket$ et $t(i) = s(i-p) + p$ pour $i \in \llbracket p+1, m+p \rrbracket$ est une injection croissante de $\llbracket 1, m+p \rrbracket$ dans $\llbracket 1, n+p \rrbracket$, vérifiant $(zx)_{t(i)} = (zy)_i$ pour tout $i \in \llbracket 1, m+p \rrbracket$, ce qui montre que $zx \sqsubseteq zy$. Pour montrer $xz \sqsubseteq yz$, on utilisera u définie par $u(i) = s(i)$ si $i \in \llbracket 1, m \rrbracket$ et $u(i) = n+i-m$ si $i \in \llbracket m+1, m+p \rrbracket$.

Question 3 • Les propriétés suivantes définissent rigoureusement la fonction $\text{esm} : \Sigma^* \times \Sigma^* \mapsto \{\mathbf{VRAI}, \mathbf{FAUX}\}$; a et b sont deux lettres.

$$\begin{aligned} \text{esm}(\varepsilon, y) &= \mathbf{VRAI} \\ \text{esm}(x, \varepsilon) &= \mathbf{FAUX} \quad \text{si } x \neq \varepsilon \\ \text{esm}(ax, ay) &= \text{esm}(x, y) \\ \text{esm}(ax, by) &= \text{esm}(ax, y) \quad \text{si } a \neq b \end{aligned}$$

• La traduction en Caml est aisée; noter que la récursivité est terminale.

```
let est_sous_mot x y =
  let m = string_length x and n = string_length y in
  let rec esm_aux i j =
    i = m or
    (j < n &
     if x.[i] = y.[j] then esm_aux (i+1) (j+1)
     else esm_aux i (j+1) )
  in esm_aux 0 0;;
```

Question 4 • Il suffit de remarquer que chaque comparaison de caractères implique l'incrémement de j ; on effectue donc au plus n comparaisons. On peut en effectuer moins, si x est sous-mot d'un préfixe *propre* de y .

Question 5 • Un mot de longueur n possède au moins un sous-mot de longueur k pour tout $k \in \llbracket 0, n \rrbracket$, donc au total au moins $n+1$ sous-mots. Ce minimum est atteint par tout mot de la forme a^n , $a \in \Sigma$.

• À chaque sous-ensemble J de $\llbracket 1, |w| \rrbracket$, on peut associer le sous-mot x de longueur $|J|$ de w défini comme suit: $x_i = w_{s(i)}$ où $s(i)$ est le i -ième élément de J lorsque l'on énumère cet ensemble selon l'ordre croissant. La majoration de l'énoncé résulte alors du fait que $\llbracket 1, |w| \rrbracket$ compte $2^{|w|}$ sous-ensembles.

Question 6 • En énumérant en ordre lexicographique, on a :

$$S(aaaba) = \{\varepsilon, a, b, aa, ab, ba, aaa, aab, aba, aaaa, aaab, aaba, aaaba\}$$

D'où $P_{aaaba} = 1 + 2X + 3X^2 + 3X^3 + 3X^4 + X^5$.

Question 7 • Soient $i \in \llbracket 0, m \rrbracket$ et $j \in \llbracket 0, n \rrbracket$. À chaque couple (x', y') formé d'un sous-mot de longueur i de x et d'un sous-mot de longueur j de y , on peut associer un sous-mot de longueur $i + j$ de xy , à savoir le mot $x'y'$. Réciproquement, soit z un sous-mot de xy de longueur k ; soit s une application strictement croissante de $\llbracket 1, k \rrbracket$ dans $\llbracket 1, m + n \rrbracket$ telle que $z_r = (xy)_{s(r)}$ pour tout $r \in \llbracket 1, k \rrbracket$. Si $s(k) \leq m$, alors z est un sous-mot de x , on peut prendre $x' = z, y' = \varepsilon, i = k$ et $j = 0$; sinon, il existe un indice i tel que $s(i) \leq m$ et $s(i + 1) > m$, auquel cas on prend $x' = z[1..i], y' = z[i + 1..k]$ et $j = k - i$. Dans les deux cas, on a déterminé deux mots x' et y' tels que $x' \sqsubseteq x, y' \sqsubseteq y$ et $z = x'y'$. Ceci montre que l'application

$$(x', y') \in \bigcup_{i+j=k} (S(x) \cap \Sigma^i) \times (S(y) \cap \Sigma^j) \mapsto x'y' \in (S(xy) \cap \Sigma^k)$$

est une surjection. Donc $[X^k]P_{xy} \leq \sum_{i+j=k} [X^i]P_x \cdot [X^j]P_y$, et par suite $P_{xy} \preceq P_x P_y$.

Question 8 • Les sous-mots de aw qui commencent par $b \neq a$ sont exactement les sous-mots de w qui commencent par b , d'où $P_{aw}^b = P_w^b$.

• On note que $[X^0]P_{aw}^a = [X^0]X(P_w^a + P_w^b + 1) = 0$ et $[X^1]P_{aw}^a = [X^1]X(P_w^a + P_w^b + 1) = 1$ (l'unique sous-mot concerné est a); si $k > |aw|$, $[X^k]P_{aw}^a = [X^k]X(P_w^a + P_w^b + 1) = 0$. Enfin, supposons $2 \leq k \leq |aw|$; $[X^k]P_{aw}^a$ est le nombre de sous-mots de aw de longueur k , commençant par a . Ils peuvent tous être obtenus en concaténant derrière l'initiale a un sous-mot de longueur $k - 1$ de w ; ceux-ci sont au nombre de $[X^{k-1}](P_w^a + P_w^b)$, d'où la contribution $[X^k]X(P_w^a + P_w^b)$; par ailleurs, $[X^k]X = 0$ puisque $k \geq 2$.

• Évidemment, $P_\varepsilon^a = 0$.

Question 9 • Par raison de symétrie, $P_{\alpha_n}^a = P_{\beta_n}^b$ et $P_{\beta_{n+1}}^a = P_{\alpha_{n+1}}^b$. Avec la question précédente, $P_{\alpha_{n+1}}^b = P_{\beta_n}^b = P_{\beta_n}^b$. Enfin, établissons $P_{\alpha_n}^a = Q_n$ par récurrence sur n ; pour $n = 0$, les deux membres sont nuls; pour $n = 1$, il vient $P_{\alpha_1}^a = P_a^a = X = Q_1$. Supposons l'égalité acquise jusqu'au rang $n + 1$ inclus; alors

$$P_{\alpha_{n+2}}^a = P_{a\beta_{n+1}}^a = X(P_{\beta_{n+1}}^a + P_{\beta_{n+1}}^b + 1) = X(P_{\alpha_n}^a + P_{\alpha_{n+1}}^a + 1) = Q_{n+2}$$

Question 10 • Rappelons les premiers termes de la suite de FIBONACCI: $F_2 = 1, F_3 = 2, F_4 = 3, F_5 = 5, F_6 = 8$ et $F_7 = 13$. Observons les premiers termes: $s(\alpha_0) = s(\varepsilon) = 1 = F_3 - 1$; $s(\alpha_1) = s(a) = 2 = F_4 - 1$; $s(\alpha_2) = s(ab) = 4 = F_5 - 1$; $s(\alpha_3) = s(aba) = 7 = F_6 - 1$ et $s(\alpha_4) = s(abab) = 12 = F_7 - 1$. Nous allons établir $s(\alpha_n) = F_{n+3} - 1$ par récurrence sur n .

$s(w) = \sum_{k=0}^{|w|} [X^k]P_w = P_w(1)$, en particulier $s(\alpha_n) = P_{\alpha_n}(1)$. Pour $n \geq 1$, en distinguant les sous-mots de α_n

qui commencent par a , ceux qui commencent par b , et le mot vide, on a $s(\alpha_n) = P_{\alpha_n}^a(1) + P_{\alpha_n}^b(1) + 1 = Q_n(1) + Q_{n-1}(1) + 1 = Q_{n+1}(1)$.

Nous nous ramenons donc à établir $Q_{n+1}(1) = F_{n+3} - 1$ pour $n \geq 1$. On note que $Q_1(1) = 1 = F_3 - 1$; et que $Q_2 = X(Q_1 + Q_0 + 1) = X^2 + X$, donc $Q_2(1) = 2 = F_4 - 1$. Supposons la relation $Q_k(1) = F_{k+2} - 1$ acquise jusqu'au rang $n + 1$ inclus; alors $Q_{n+2}(1) = Q_{n+1}(1) + Q_n(1) + 1 = F_{n+4} - 1 + F_{n+3} - 1 + 1 = F_{n+5} - 1$, ce qui établit l'assertion au rang $n + 1$.

Question 11 • Il est clair que $\alpha_n \sqsubseteq \alpha_{n+1}$ et $\beta_n \sqsubseteq \beta_{n+1}$. On a $[X^{n+1}]Q_{n+1} = 1$, tandis que $[X^{n+1}]Q_n = 0$. Pour $k > n + 1$, $[X^k]Q_n = [X^k]Q_{n+1} = 0$. Enfin, pour $k \leq n$, $[X^k]Q_n$ est le nombre de sous-mots commençant par a et de longueur k de α_n , qui est lui-même sous-mot de α_{n+1} ; donc $[X^k]Q_n$ est au plus égal à $[X^k]Q_{n+1}$. Ainsi, $Q_n \prec Q_{n+1}$.

Question 12 • Soit u de longueur n , distinct de α_n et de β_n : u contient donc au moins un facteur de la forme aa ou bb , disons aa pour fixer les idées. En observant la dernière occurrence dans u de ce facteur, on constate que u peut s'écrire $u = xa\alpha_p$, avec $0 < p < n$. Considérons alors $v = x\alpha_{p+1}$: $|v| = n$. On a $P_{a\alpha_p}^a = X(P_{\alpha_p}^a + P_{\alpha_p}^b + 1) = X(Q_p + Q_{p-1} + 1) = Q_{p+1} = P_{\alpha_{p+1}}^a$ et $P_{a\alpha_p}^b = P_{\alpha_p}^b = Q_{p-1} < Q_p = P_{\alpha_{p+1}}^b$. On en déduit $P_u^a \preceq P_v^a$ et $P_u^b \prec P_v^b$. Par suite $P_u = P_u^a + P_u^b + 1 \prec P_v^a + P_v^b + 1 = P_v$.

Question 13 • Résumons: $\max_{w \in \Sigma^n} s(w) = F_{n+3} - 1$, le maximum étant atteint par les mots α_n et β_n , et eux seuls.

FIN

[plsmc1cor]

Version du 23 février 2008