Option Informatique en Spé MP et MP*
Nombre maximal de sous-mots d'un mot

► Σ désigne un alphabet contenant au moins deux lettres a et b. Un mot x est un *sous-mot* d'un mot y (ce que l'on notera $x \sqsubseteq y$) si, notant $m = |x|$ et $n = |y|$, il existe une injection croissante s de $[\![1, m]\!]$ dans $[\![1, n]\!]$ telle que $x_i = y_{s(i)}$ pour tout $i \in [\![1, m]\!]$. Ceci revient à dire que x peut se déduire de y en *supprimant* certains caractères ; ou, inversement, que y peut se déduire de x en *insérant* des caractères. Ainsi, *corne* est un sous-mot de *re<u>co</u>uv<u>r</u>a<u>n</u>t<u>e</u>*.

► On notera que, si x est facteur de y, alors x est un sous-mot de y ; mais la réciproque n'est pas vraie.

► On note $|X|$ le cardinal de l'ensemble fini X, et $\lg x$ le logarithme en base 2 du réel $x > 0$.

Question 1 • Citez quelques sous-mots remarquables du mot *découvrable*.

Question 2 • Justifiez rapidement : la relation $\sqsubseteq$ est un ordre, compatible avec la concaténation : si $x \sqsubseteq y$, alors $xz \sqsubseteq yz$ et $zx \sqsubseteq zy$.

Question 3 • Rédigez en Caml une fonction :

```
est_sous_mot : string -> string -> bool
```

spécifiée comme suit : `est_sous_mot x y` dit si x est un sous-mot de y. Vous justifierez en toute rigueur la validité de cette fonction.

Question 4 • Montrez qu'avec cette fonction, le coût de `est_sous_mot x y` (exprimé en nombre de comparaisons de caractères) est au plus égal à n.

► Soit $w \in \Sigma^*$; on note $S(w)$ l'ensemble des sous-mots de w et $s(w) = \big|S(w)\big|$.

Question 5 • Soit $n \in \mathbb{N}$; calculez $\min\{s(w) \mid w \in \Sigma^n\}$ et précisez les mots pour lesquels ce minimum est atteint. Justifiez la majoration $s(w) \leqslant 2^{|w|}$.

► Dans la suite de cette partie, on suppose que Σ se réduit à deux lettres, donc $\Sigma = \{a, b\}$. On se propose d'expliciter $\max\{s(w) \mid w \in \Sigma^n\}$, en précisant les mots pour lesquels ce maximum est atteint.

► Soit $P \in \mathbb{R}[X]$; pour $k \in \mathbb{N}$, on note $[X^k]P$ le coefficient de X^k dans P, ainsi $P = \sum\limits_{k \in \mathbb{N}} \big([X^k]P\big)X^k$. À un langage fini L, on associe le polynôme P_L défini par :

$$P_L = \sum_{k \in \mathbb{N}} \big|L \cap \Sigma^k\big| X^k$$

$[X^k]P_L$ est donc le nombre de mots de longueur k qui appartiennent à L.

► Soit $w \in \Sigma^*$; on note $P_w = P_{S(w)}$, et $P_w^a = P_{S(w) \cap a\Sigma^*}$; ainsi $[X^k]P_w^a$ est le nombre de sous-mots de w qui commencent par la lettre a et dont la longueur est k.

► Soient U et V deux polynômes ; on note $U \preceq V$ si $[X^k]U \leqslant [X^k]V$ pour tout $k \in \mathbb{N}$; on note $U \prec V$ si $U \preceq V$ et $U \neq V$. Clairement, $\preceq$ est un ordre partiel, compatible avec l'addition.

Question 6 • Explicitez P_{aaaba}.

Question 7 • Établissez la relation $P_{xy} \preceq P_x P_y$.

Question 8 • Justifiez $P_{aw}^b = P_w^b$, puis $P_{aw}^a = X(P_w^a + P_w^b + 1)$. Combien vaut P_ε^a ?

► On définit une suite $(Q_n)_{n \in \mathbb{N}}$ de polynômes par les formules $Q_0 = 0$, $Q_1 = X$ et $Q_{n+2} = X(Q_{n+1} + Q_n + 1)$. On définit également deux suites $(\alpha_n)_{n \in \mathbb{N}}$ et $(\beta_n)_{n \in \mathbb{N}}$ de mots par les formules $\alpha_0 = \beta_0 = \varepsilon$, $\alpha_{n+1} = a\beta_n$ et $\beta_{n+1} = b\alpha_n$. Ainsi $|\alpha_n| = |\beta_n| = n$, $\alpha_n = ababa\ldots$ et $\beta_n = babab\ldots$; plus précisément, $\alpha_{2p} = (ab)^p$, $\alpha_{2p+1} = (ab)^p a$, $\beta_{2p} = (ba)^p$ et $\beta_{2p+1} = (ba)^p b$.

Question 9 • Justifiez : $P_{\alpha_n}^a = P_{\beta_n}^b = P_{\beta_{n+1}}^a = P_{\alpha_{n+1}}^b = Q_n$.

Question 10 • La *suite de Fibonacci* est définie par les relations $F_0 = 0$, $F_1 = 1$ et $F_{n+2} = F_{n+1} + F_n$. Explicitez $s(\alpha_n)$ en fonction d'un ou plusieurs termes de cette suite.

Question 11 • Justifiez la relation $Q_n \prec Q_{n+1}$.

Question 12 ⋆⋆⋆ • Soit u un mot de longueur n, distinct de α_n et β_n. Montrez qu'il existe un mot v de même longueur, et vérifiant $P_u \prec P_v$.

Question 13 • Et maintenant, concluez !

FIN