Option Informatique en Spé MP et MP*
Plus long sous-mot commun : ce qu'il faut savoir

► Σ désigne un alphabet contenant au moins deux lettres a et b.

Définition : 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 vraie si $|\Sigma| = 1$.

Question : de quelle célébrité le nom apparaît-il comme sous-mot du mot *découvrable* ?

Proposition : la relation $\sqsubseteq$ est un ordre, compatible avec la concaténation : si $x \sqsubseteq y$, alors $xz \sqsubseteq yz$ et $zx \sqsubseteq zy$.

Preuve

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

• Antisymétrie : soient x et y vérifiant $x \sqsubseteq y$ et $y \sqsubseteq x$. Soient s une injection croissante de $[\![1, m]\!]$ dans $[\![1, n]\!]$ telle que $y_{s(i)} = x_i$ pour tout $i \in [\![1, m]\!]$, et t une injection croissante de $[\![1, n]\!]$ dans $[\![1, m]\!]$ telle que $x_{t(j)} = y_j$ pour tout $j \in [\![1, n]\!]$. On a nécessairement $n = m$; puis s, injection croissante de $[\![1, m]\!]$ 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 $[\![1, m]\!]$ dans $[\![1, n]\!]$ telle que $y_{s(i)} = x_i$ pour tout $i \in [\![1, m]\!]$, et t une injection croissante de $[\![1, n]\!]$ dans $[\![1, p]\!]$ telle que $z_{t(j)} = y_j$ pour tout $j \in [\![1, n]\!]$. Alors $z_{(t \circ s)(i)} = x_i$ pour tout $i \in [\![1, m]\!]$, 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 $[\![1, m]\!]$ dans $[\![1, n]\!]$ adaptée. Soit z de longueur p ; alors $t : [\![1, m+p]\!] \mapsto [\![1, n+p]\!]$ définie par $t(i) = i$ pour $i \in [\![1, p]\!]$ et $t(i) = s(i-p) + p$ pour $i \in [\![p+1, m+p]\!]$ est une injection croissante de $[\![1, m+p]\!]$ dans $[\![1, n+p]\!]$, vérifiant $(zx)_{t(i)} = (zy)_i$ pour tout $i \in [\![1, m+p]\!]$, ce qui montre que $zx \sqsubseteq zy$. Pour montrer $xz \sqsubseteq yz$, on utilisera u définie par $u(i) = s(i)$ si $i \in [\![1, m]\!]$ et $u(i) = i - m$ si $i \in [\![m+1, m+p]\!]$.

► Par programmation dynamique, on peut trouver un *plus long sous-mot commun* à deux mots u et v pour un coût $\mathcal{O}\big(|u| \times |v|\big)$. En revanche, la recherche d'un mot v de longueur minimale et admettant chacun des mots $u_1, \ldots, u_n$ comme sous-mots est un problème NP-complet, n'admettant même pas d'algorithme d'approximation.

Définition : une *antichaîne* d'un ensemble ordonné $(E, \preceq)$ est un ensemble d'éléments de E deux à deux incomparables.

Exemples : dans $\mathcal{P}(E)$ muni de l'inclusion, les singletons forment une anti-chaîne. Dans $\mathbb{N}$ muni de la relation «divise», les nombres premiers forment une anti-chaîne. Dans un langage L, l'ensemble des mots de longueur minimale forment une anti-chaîne pour la relation $\sqsubseteq$ comme pour la relation «est facteur de».

Théorème (Higman) : il n'existe pas d'antichaîne infinie pour l'ordre $\sqsubseteq$ sur un alphabet fini.

Preuve : voir le *Handbook of Formal Languages.*

FIN