Le théorème du sous-mot

Référence : Bruno PETAZZONI, 16 problèmes d'informatique

2011-2012

On définit la relation $<$ sur les mots par : si $u = u_1 \dots u_n$ et $v = v_1 \dots v_m$, on a $u < v$ *si et seulement si* il existe une application strictement croissante f de $[1, n]$ sur $[1, m]$ telle que $u_i = v_{f(i)}$.

Pour un mot w et un langage L, on notera L_w l'ensemble des sur-mots de w, *i.e*

$$L_w = \{v \in \Sigma^* \mid w < v\}.$$

Pour un langage L, on notera $\hat{L}$ l'ensemble des sur-mots des mots de L, et $SM(L)$ l'ensemble des sous-mots des mots de L.

Enfin, on appellera *antichaîne* sur Σ un langage L dont les mots ne sont pas comparables deux à deux pour $<$.

Théorème 1 : du sous-mot
Si $L \in \mathfrak{p}(\Sigma^*)$, alors $SM(L)$ est rationnel.

Démonstration. La démonstration repose sur le :

Lemme 2 : de Higman
Toute antichaîne sur Σ pour $<$ est finie.

On se fixe un langage L quelconque.

Lemme 3
$\hat{L}$ est rationnel.

Démonstration. On considère $L_\downarrow$ l'ensemble des mots de L minimaux pour $<$. Alors nécessairement, $L_\downarrow$ est une antichaîne, donc est finie.

Pour tout mot u de $\hat{L}$, il existe un mot $w \in L_\downarrow$ tel que $w < u$, et donc finalement :

$$\hat{L} = \bigcup_{w \in L_\downarrow} L_w.$$

Chaque L_w est rationnel ($L_w = \Sigma^* w_1 \Sigma^* \cdots \Sigma^* w_n \Sigma^*$), et donc par stabilité des langages rationnels par union finie, on a $\hat{L}$ rationnel. $\Diamond$

Maintenant, considérons $K = \Sigma^* \backslash SM(L)$. On va montrer que $K = \hat{K}$, et on aura alors la rationnalité de K, et donc de $SM(L)$ par clotûre des langages rationnels par complémentarisation.

L'inclusion $K \subseteq \hat{K}$ est triviale.

Soit donc $x \in \hat{K}$. Il existe $w \in K$ tel que $w < x$. Alors si $x \in SM(L)$, alors w aussi, ce qui est une contradiction. D'où le résultat. $\square$

Montrons le lemme de Higman :

Démonstration. On suppose que A est une antichaîne infinie sur Σ. On note alors

$$q = |\Sigma \text{ et } n = \min_{u \in A} |u|.$$

On suppose de plus que q et n sont minimaux : les antichaînes sur un alphabet plus petit que Σ sont finies, et celles dont le mot de longueur minimale est plus court que n aussi.

1. On remarque que $q > 1$ et $n > 1$

   Démonstration. Si $q = 1$, alors toute antichaîne ne contient qu'un élément.
   Si $n = 1$, alors u de longueur 1 est une lettre, et donc $A \backslash \{u\}$ antichaîne infinie sur $\Sigma \backslash \{u\}$ de taille $q - 1 \leftarrow$ contradiction avec la minimalité de q. $\Diamond$

2. On fixe u de longueur n. On définit A' comme l'ensemble des mots de A qui sont sur-mots de $u[1, n-1]$. On a alors $A \backslash A'$ fini, donc A' infini.

   Démonstration. Le langage $(A \backslash A') \cup \{u[1, n-1]\}$ est une antichaîne sur Σ, avec un mot minimal de longueur $n - 1$. Par minimalité de n, cette antichaîne est finie. $\Diamond$

3. On énumère A' : $A' = \{v_1, v_2, \ldots, v_i \ldots\}$, et on écrit chaque v_i comme

   $$v_i = z_{i,1} u_1 z_{i,2} u_2 \cdots z_{i,n-1} u_{n-1} z_{i,n},$$

   où pour $j \leqslant n - 1$, $z_{i,j} \in (\Sigma \backslash \{u_j\})^*$.
   Alors $z_{i,n} \in (\Sigma \backslash \{u_n\})^*$.

   Démonstration. Sinon, on retrouve u dans v_i, et donc $u < v_i$ ce qui est impossible par définition d'antichaîne. $\Diamond$

4. On définit $Z_j = \{z_{i,j} \mid i \in \mathbb{N}\}$.
   Alors Z_j a un nombre fini d'éléments minimaux pour $<$.

   Démonstration. L'ensemble des éléments minimaux de Z_j est une antichaîne, sur un alphabet de cardinale plus petit strictement que q, donc est fini. $\Diamond$

5. Si Z_j est infini, alors à extraction près, on a pour tout j :

   $$z_{i,j} < z_{i+1,j}.$$

   Démonstration. Soit $s(1)$ le plus petit indice tel qu'il n'y ait plus d'éléments minimaux : $z_{s(1),j}$ n'est pas minimal, et donc $\exists s(2) > s(1)$ tel que $z_{s(1),j} < z_{s(2),j}$.
   En itérant le procédé, on construit la fonction s sur $\mathbb{N}$. $\Diamond$

6. L'un au moins des Z_j est infini.

   Démonstration. On a $A' = Z_1 u_1 Z_2 u_2 \cdots Z_{n-1} u_{n-1} Z_n$, et A' infini. $\Diamond$

7. En applicant les extractions correspondant aux Z_j infinis, on va se retrouver avec des mots tous comparables entre eux. D'où une contradiction avec le fait que A' est une antichaîne. D'où le résultat.

$\square$