

Algorithmique sur les automates — 23 Février 98

Les trois exercices et le problème sont indépendants. Dans *tout l'énoncé*, l'alphabet A est fixé. On rappelle qu'un bord d'un mot $x \neq \varepsilon$ est un mot distinct de x à la fois préfixe et suffixe de x , et que le bord de x est le plus long bord de x . On écrira $x \preceq_p y$ (resp. $x \preceq_s y$) si x est un préfixe (resp. un suffixe) de y .

Exercice 1 Deux mots x et y sont *conjugués* s'il existe $u, v \in A^*$ tels que $x = uv$, $y = vu$. Montrer que l'on peut déterminer en temps $O(n)$ si deux mots de longueur n sont conjugués.

Exercice 2 Soit $x \in A^*$. Un facteur y de x est *répété* s'il a deux occurrences distinctes dans x .

1 À l'aide de l'arbre compact des suffixes, montrer que l'on peut déterminer le plus long facteur répété d'un mot x en temps $O(|x|)$.

2 Soient x_1 et x_2 deux mots de A^* . Montrer, en considérant le mot $x_1\#x_2\$$ (où $\#$ et $\$$ ne sont pas dans A) que l'on peut déterminer en temps $O(|x_1| + |x_2|)$ le plus long facteur commun à x_1 et x_2 . Généraliser à un nombre quelconque (fixé) de mots.

Exercice 3 Un *carré* dans x est un facteur de x de la forme yy , avec $y \in A^*$. Le mot x est dit *sans carré* si le seul carré dans x est ε .

1 Soient x et y deux mots sans carré. Montrer que xy contient un carré non vide si et seulement si l'une des deux conditions suivantes est vraie :

$$\exists z_1, z_2 \in A^*, \quad z_1 \neq \varepsilon, \quad z_1 \preceq_s x \text{ et } z_2 z_1 z_2 \preceq_p y \quad (1)$$

$$\exists z_1, z_2 \in A^*, \quad z_2 \neq \varepsilon, \quad z_1 z_2 z_1 \preceq_s x \text{ et } z_2 \preceq_p y \quad (2)$$

On dit que $d(x, y)$ (resp. $g(x, y)$) est vraie si x et y sont sans carré et satisfont (1) (resp. satisfont (2)).

2 On suppose que $d(x, y)$ est vraie. Soit uvu le plus court préfixe de y tel que $v \neq \varepsilon$ soit suffixe de x . Montrer que u est le bord de uvu , et que v est le plus long suffixe commun à x et uv .

3 Mêmes notations et hypothèses que dans la question précédente. On admet que si $ztwz$ est un préfixe de y tel que $|ztwz| > |uvu|$ et tel que w soit le plus long suffixe commun à x et à ztw , alors $|uv| < \max(|zt|, \frac{1}{2}|ztw|)$. On suppose dans un premier temps que le bord de chaque préfixe de y est connu. Montrer que l'on peut déterminer si $d(x, y)$ est vraie en temps $O(|y|)$ en lisant y de droite à gauche.

4 En déduire que si x et y sont sans carré, on peut tester en temps $O(|xy|)$ si xy est sans carré.

Problème On note, pour tout mot $x \in A^*$ et tout entier $n \geq 1$:

- $P_n(x)$ l'ensemble des préfixes de x de longueur au plus $n - 1$.
- $S_n(x)$ l'ensemble des suffixes de x de longueur au plus $n - 1$.
- $F_n(x)$ l'ensemble des facteurs de x de longueur exactement n .

1 Pour un entier $n \geq 1$, on considère la relation \approx_n définie sur A^* par :

$$x \approx_n y \iff (P_n(x) = P_n(y), \quad S_n(x) = S_n(y), \quad F_n(x) = F_n(y))$$

1.a Montrer que \approx_n est une congruence.

1.b Montrer que \approx_n est d'indice fini. Comparer \approx_n et \approx_{n+1} .

2 Un langage est dit *localement testable* s'il est union de \approx_n -classes pour un certain $n \geq 1$. Montrer qu'un langage localement testable est rationnel.

2.a Plus précisément, montrer qu'un langage est localement testable s'il est dans l'algèbre de Boole engendrée par les langages de la forme uA^* , A^*u , A^*uA^* où $u \in A^*$.

2.b Que peut-on dire du monoïde syntaxique d'un tel langage? Le langage $(a^2)^*$ sur l'alphabet $\{a\}$ est-il localement testable? Même question pour le langage B^* , où $B \subseteq A$, puis pour le langage $(ab)^*$ sur l'alphabet $\{a, b\}$.

3 Soit L un langage localement testable, \equiv_L la congruence syntaxique de L , et u, v, w tels que

$$u \neq \varepsilon \text{ et } u \equiv_L u^2 \quad (3)$$

$$v, w \in A^*u \cap uA^* \quad (4)$$

Montrer que pour tout $k \geq 1$, $v^2 \equiv_L vu^k v$ et $vw \equiv_L vu^k w$. En déduire que

$$v \equiv_L v^2 \text{ et } vw \equiv_L wv \quad (5)$$

4.a Le langage $(ab \cup ba)^*$ est-il sans étoile?

4.b Montrer que le langage $(ab \cup ba)^*$ n'est pas localement testable.

5 Montrer qu'étant donné un langage L , il existe un algorithme pour tester si (5) est vérifiée pour *tous* les triplets $(u, v, w) \in (A^+)^3$ qui satisfont (3) et (4).

On veut maintenant montrer que réciproquement, si (5) est vérifiée pour tous les triplets $(u, v, w) \in (A^+)^3$ qui satisfont (3) et (4), alors L est localement testable. On établit d'abord dans la question **6** un résultat intermédiaire, utilisé en **7** et **8** (ces deux dernières questions peuvent bien sûr être traitées en admettant **6**).

6 Soit \mathcal{A} un automate *déterministe* et Q son ensemble d'états. Pour $p \in Q$ et $x \in A^*$, on note $p \cdot x$ l'état atteint en lisant x à partir de p , s'il existe. Le chemin étiqueté x et partant de p est noté $p \xrightarrow{x} p \cdot x$. On munit \mathcal{A} d'une fonction de sortie de $Q \times A$ dans un monoïde M : $(p, a) \mapsto p * a \in M$. On étend cette fonction à $Q \times A^*$ en posant $p * \varepsilon = 1_M$ et $p * (xa) = (p * x)((p \cdot x) * a)$. On suppose dans toute cette question que

$$\forall p \in Q, \quad \forall x, y \in A^*, \quad (p \cdot x = p \cdot y = p) \implies \begin{cases} p * x^2 = p * x \\ p * xy = p * yx \end{cases} \quad (6)$$

6.a Soient $p, q, r \in Q$ et $x, y \in A^*$ tels que

$$- p \cdot x = q, \quad q \cdot y = r,$$

- le chemin $p \xrightarrow{x} q$ contient toutes les transitions du chemin $q \xrightarrow{y} r$.

On veut montrer qu'il existe une factorisation $x = x_1 x_2$ de x telle que $p \cdot x_1 = r$ et $p * x = p * (x_1 x_2)$. Vérifier d'abord le résultat pour $y = \varepsilon$. On pose ensuite $y = za$, $a \in A$. Montrer qu'il existe une factorisation $x = x' a x''$ de x telle que $p \cdot x' = q \cdot z$ et qui vérifie $p * x = p * x' y x''$. Conclure.

6.b Avec les mêmes notations qu'en **6.a**, et toujours en supposant que le chemin $p \xrightarrow{x} q$ contient toutes les transitions du chemin $q \xrightarrow{y} r$, en déduire que si $q = r$, alors $p * xy = p * x$.

Soient $p, q \in Q$ et $x, y \in A^*$. On suppose que $p \cdot x = p \cdot y$ et que les chemins $p \xrightarrow{x} q$ et $p \xrightarrow{y} q$ utilisent exactement les mêmes transitions. On veut montrer qu'alors

$$p * x = p * y \quad (7)$$

Quitte à supprimer des transitions, on suppose que $p \xrightarrow{x} q$ et $p \xrightarrow{y} q$ utilisent exactement toutes les transitions de \mathcal{A} .

6.c Déduire de **6.b** que s'il existe un mot z tel que $q \cdot z = p$, alors $p * x = p * y$.

6.d Montrer que si un tel mot n'existe pas, il existe des factorisations $x = x_0 a x_1$ et $y = y_0 a y_1$ telles que $p \cdot x_0 = p \cdot y_0$ et telles que

- tous les états traversés sur le chemin d'étiquette x_0 (resp. y_0) ne sont pas accessibles de q ,
- tous les états traversés sur le chemin d'étiquette x_1 (resp. y_1) sont accessibles de q ,

Montrer que les chemins ci-dessus d'étiquettes x_0 et y_0 utilisent les mêmes transitions. Même question pour les chemins d'étiquettes x_1 et y_1 . Montrer que ces ensembles de transitions sont inclus dans l'ensemble des transitions utilisés par les chemins d'étiquette x et y . Conclure.

7 Soit $L \subseteq A^*$ rationnel. Soit M le monoïde syntaxique de M et soit $n > \text{Card}(M)$. Soit enfin $\varphi_L : A^* \rightarrow M$ le morphisme syntaxique. On considère l'automate avec sortie \mathcal{A}

- dont les états sont les mots de n lettres,
- dont la fonction de transition est donnée par $(a_1 \cdots a_n) \cdot a = a_2 \cdots a_n a$,
- dont la fonction de sortie est donnée par $p * a = \varphi(pa)$.

7.a En considérant la suite de ses préfixes, montrer que tout état p de \mathcal{A} admet une factorisation $p = p_1 p_2 p_3$ satisfaisant $p_2 \neq \varepsilon$ et $\varphi(p_1 p_2) = \varphi(p_1)$.

7.b Soit x un mot qui étiquette un chemin de p à p . Montrer que p_3 est un bord de $p_3 x$. Soit x_1 tel que $p_3 x = x_1 p_3$. Montrer que p_2 est un bord de $p_2 x_1$, et soit x_2 tel que $p_2 x_1 = x_2 p_2$. Montrer enfin que p_1 est un bord de $p_1 x_2$, et que $\varphi(p_1 p_2 x_1) = \varphi(p_1 x_2)$.

7.c Soit ℓ un entier tel que $\varphi(p_2^\ell)$ soit idempotent. Montrer que $\varphi(p x) = \varphi(p_1 p_2^\ell x_2 p_2^\ell p_3)$. On suppose (5) vérifiée pour tous les triplets $(u, v, w) \in (A^+)^3$ qui satisfont (3) et (4). Montrer que $\varphi(p x) = \varphi(p x^2)$. De même, soit y un mot étiquetant un chemin de p à p . Montrer que $\varphi(p x y) = \varphi(p y x)$. En déduire que l'automate \mathcal{A} satisfait (6), et donc (7) d'après **6**.

8 On suppose toujours (5) vérifiée pour tous les triplets (u, v, w) qui satisfont (3) et (4).

8.a Soient x et y deux mots de longueur au moins n tels que $x \approx_n y$, soit p (resp. q) leur préfixe commun de longueur n (resp. leur suffixe commun de longueur n). On pose $x = p x'$, $y = p y'$. Montrer que $p \xrightarrow{x'} q$ et $p \xrightarrow{y'} q$ sont deux chemins de \mathcal{A} utilisant les mêmes transitions.

8.b En utilisant (7), en déduire que $\varphi(x) = \varphi(y)$ et que L est localement testable.