Info Spés - 23 Octobre 2001 **DS 1 - 3 Heures**

Langages et automates

Ce sujet est constitué de deux problèmes. Le premier concerne les *codes*, ensembles de mots "libres" au sens du monoïde $(A^*, .)$. Le second concerne les langages rationnels : on y montre que la classe des langages rationnels est stable par certaines applications.

Ces deux exercices ont été largement pompés sur des feuilles d'exercices de Vincent Bouchitté : `http://www.ens-lyon.fr/~vbouchit` : en particulier, je ne m'étais jamais intéressé aux langages $\frac{1}{2}L$ et $\frac{2}{3}L$.

1 Codes

Soit A un alphabet. Un *code* sur A est un ensemble de mots C tel que tout mot de C^+ se décompose de façon unique $w_1...w_k$, avec tous les $w_i \in C$. En d'autres termes, si $w_1...w_k = w'_1...w'_j$ avec chaque w_i et chaque w'_i dans C, alors $j = k$ et $w_i = w'_i$ pour tout $i \in [\![1, k]\!]$.

1. Parmi les langages suivants déterminer ceux qui sont des codes[1] :
   - $L_1 = \{a, b\}$.
   - $L_2 = ab^*$.
   - $L_3 = \{a, ab, ba\}$.
   - $L_4 = \{a, ba, bb\}$.
   - $L_5 = \{aa, baa, ba\}$.
   - $L_6 = \{a, abbba, babab, bb\}$.
   - $L_7 = a^+b^+$.
   - $L_8 = a^+b^*$.
   - $L_9 = \{a^nb^n \mid n \in \mathbb{N}^*\}$.
2. Soit $w \in A^*$. Montrer que w est un code si et seulement si $w \neq \varepsilon$.
3. Soit X une partie de A^* ne contenant pas ε et telle qu'aucun mot de X n'est préfixe d'un autre mot que lui-même. Montrer que X est un code.
4. Même chose en remplaçant "préfixe" par "suffixe".
5. Soient $w_1, w_2 \in A^*$ avec $w_1 \neq w_2$. Montrer que les trois propositions suivantes sont équivalentes :
   (a) $\{w_1, w_2\}$ est un code ;
   (b) $w_1w_2 \neq w_2w_1$;
   (c) w_1 et w_2 ne sont pas puissances d'un même mot.

   Pour c ⇒ a, on pourra raisonner par l'absurde et considérer un contre-exemple "de taille minimale".
6. Soit L un langage rationnel. Proposer un algorithme permettant de décider si L est un code (difficile). Application : reprendre L_7 et L_8 (la construction proposée ne sera pas lue si elle n'est pas appliquée à ces langages).

[1] Bien entendu, on demande une preuve à chaque fois : une formalisation complète pouvant être pénible, on attend des arguments pertinents et convaincants

7. Dans la question précédente, évaluer la complexité des différentes constructions intervenant dans l'algorithme, par exemple par rapport à la longueur d'une expression rationnelle associée à L.

   Cette question nécessite un certain temps pour être traitée de façon convaincante : vous serez payés en fonction de la pertinence des propos...

2 Quelques constructions

L'alphabet de travail A pourra être pris égal à $\{a,b\}$ pour alléger les notations si on le souhaite (mais les résultats sont indépendants de cette hypothèse).

Si L est un langage sur A, on définit 7 autres langages associés :

- $\overline{L} = \{\widetilde{w} \mid w \in L\}$ (si $w = \alpha_1\alpha_2...\alpha_n$, alors $\widetilde{w} = \alpha_n...\alpha_2\alpha_1$).
- $Init(L) = \{u \in A^* \mid \exists v \in A^*;\ uv \in L\}$.
- $Min(L) = \{u \in L \mid$ aucun préfixe strict de u n'est dans $L\}$
- $Max(L) = \{u \in L \mid \forall v \in A^+,\ uv \notin L\}$
- $Cycle(L) = \{uv \mid u, v \in A^* \text{ et } vu \in L\}$
- $\dfrac{1}{2}L = \{u \in A^* \mid \exists v \in A^*; |u| = |v| \text{ et } uv \in L\}$
- $\dfrac{2}{3}L = \{u \in A^* \mid \exists v_1, v_2 \in A^*; |u| = |v_1| = |v_2| \text{ et } v_1uv_2 \in L\}$

1. Donner un automate déterministe complet reconnaissant le langage $L_1 = a^*b$.
2. Donner une expression rationnelle simple décrivant $\overline{L_1}$, ainsi qu'un automate fini déterministe reconnaissant ce langage.
3. Même chose avec $Init(L_1)$, $Min(L_1)$, ... $\dfrac{2}{3}L_1$.
4. Déterminer les 7 langages associés à $L_2 = \{a^nb^n \mid n \in \mathbb{N}\}$.
5. Montrer que si L est rationnel, alors chacun des 7 langages décrits plus haut est rationnel. Pour chaque construction, on commencera par donner l'idée maitresse avant de formaliser, et on appliquera sa construction à L_1 OBLIGATOIREMENT.

   Indications :
   - *Après l'avoir justifié (!), on pourra (...) privilégier le point de vue "reconnaissable".*
   - *Pour $Cycle(L)$ (difficile), on pourra ε-transitioner subtilement des copies d'un automate reconnaissant L...*
   - *Pour $\dfrac{1}{2}L$ (très difficile), on pourra étudier la suite d'ensemble d'états $Q_n = \{q \in Q \mid \exists w \in A^n;\ \delta(q, w) \in F\}$... avec Q, δ et F ce qu'on pense.*
   - *Pour $\dfrac{2}{3}L$ (qui est vraiment très difficile), on pourra travailler dans le même esprit que $\dfrac{1}{2}L$, en pensant en plus à l'automate qui intervenait pour reconnaître $\sqrt[N]{L}$ en TD...*
6. (S'il vous reste du temps...) Soit $L_3 = a^*b^*$: reprendre les questions 1-2-3 avec L_3, ainsi que les constructions du 5.