TD sur la théorie des langages 23 novembre 2007

Prérequis : TD 8 + transformations de grammaires, lemme de l'étoile, grammaires linéaires.
Durée : 1 h 50

TD 9 – Langages algébriques et rationnels

Problème

1 Propriétés de fermetures

Dans cette partie, on notera A un alphabet quelconque, A^* l'ensemble des mots sur A (*i.e.* le plus grand langage sur A au sens de l'inclusion) et $\mathcal{P}(A^*)$ l'ensemble des langages sur A.

Définition 1 : Soit ϕ une opération d'arité n. On rappelle qu'un ensemble E est **fermé** (**stable**) par ϕ ssi :

$$L_1,\ L_2,\ \ldots,\ L_n \in E \quad \Rightarrow \quad \phi(L_1, L_2, \ldots, L_n) \in E$$

Remarque 1 : On sait que, par définition, l'ensemble des langages rationnels $Rat(A^*)$ est stable par concaténation, union et itéré. On cherche alors d'autres propriétés de fermeture de $Rat(A^*)$. Pour cela, on s'appuiera sur le résultat suivant (théorème de Kleene) :

Théorème 1 : Un langage est rationnel ssi il est reconnaissable par un automate fini.

Remarque 2 : On en déduit que l'ensemble $Rat(A^*)$ est stable pour l'opération ϕ ssi :

$$\begin{gathered} L_1,\ L_2,\ \ldots,\ L_n \text{ sont reconnaissables par des automates finis } \mathcal{A}_1,\ \mathcal{A}_2,\ \ldots,\ \mathcal{A}_n \\ \Rightarrow \\ \phi(L_1, L_2, \ldots, L_n) \text{ est reconnaissable par un automate fini } \mathcal{A}_\phi \end{gathered}$$

Il suffit alors de construire $\mathcal{A}_\phi$ à partir de $\mathcal{A}_1,\ \mathcal{A}_2,\ \ldots,\ \mathcal{A}_n$.

Question 1 : Montrer que $Rat(A^*)$ est fermé par complémentaire. En déduire (en utilisant les lois de De Morgan) que $Rat(A^*)$ est fermé par intersection.

Question 2 : Montrer que $Rat(A^*)$ est stable par l'opération miroir.

Remarque 3 : On peut montrer de la même façon que $Rat(A^*)$ est fermé par préfixe, suffixe, facteur, quotient gauche et quotient droit.

On cherche désormais à déterminer les propriétés de fermeture de l'ensemble des langages algébriques. On s'appuiera sur la propriété suivante :

Théorème 2 : Un langage est algébrique ssi il est engendré par une grammaire algébrique.

Remarque 4 : Pour montrer que $Alg(A^*)$ est stable pour une opération ϕ, il suffit alors de considérer n grammaires algébriques $G_1,\ G_2,\ \ldots,\ G_n$ engendrant respectivement $L_1,\ L_2,\ \ldots,\ L_n$ et de construire une grammaire algébrique engendrant $\phi(L_1, L_2, \ldots, L_n)$ à partir de $G_1,\ G_2,\ \ldots,\ G_n$.

Question 3 : Montrer que $Alg(A^*)$ est fermé par union, concaténation et itération.

Remarque 5 : $Alg(A^*)$ n'est pas fermé par intersection ni par complémentaire. En revanche, $Alg(A^*)$ est fermé par intersection avec un langage rationnel.

2 Langages algébriques non rationnels

On note $nbOcc : A^* \times A \to \mathbb{N}$ l'application qui à un mot α et une lettre a associe le nombre d'occurences de a dans α.

Question 4 : Soit $L_1 = \{\alpha \in A^* \mid nbOcc(\alpha, a) = nbOcc(\alpha, b)\}$. Donner une grammaire algébrique G_1 engendrant L_1 et démontrer par induction structurelle que $\mathcal{L}(G_1) = L_1$.

Question 5 : Soit la grammaire linéaire $G_2 = \left(\{A, B\}, \{a, b\}, \left\{ \begin{array}{ccc|c} A & \to & aA & B \\ B & \to & bB & \varepsilon \end{array} \right\}, A\right)$. Donner un automate équivalent à (*i.e.* reconnaissant le langage engendré par) G_2.

Question 6 : Déterminiser l'automate obtenu à la question précédente.

Question 7 : Déterminer une expression rationnelle dénotant $L_2 = \mathcal{L}(G_2)$.

Question 8 : Montrer, en utilisant le lemme de l'étoile, que le langage $L_3 = \{a^n b^n \mid n \in \mathbb{N}\}$ n'est pas rationnel.

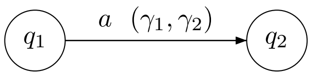
Question 9 : Trouver le langage $L = L_1 \cap L_2$. En déduire que L_1 n'est pas rationnel.

3 À la découverte des automates à piles

Définition 2 : On appelle **automate à pile** tout 6-uplet $\mathcal{A} = (A, Q, q_0, R, F, P)$ où
- A est l'**alphabet d'entrée**
- Q est un ensemble fini, l'ensemble des **états**
- $q_0 \in Q$ est l'**état initial**
- $F \subset Q$ est l'ensemble des **états finaux** (**terminaux**)
- Γ est l'**alphabet de la pile**
- $R \subset Q \times (P \cup \varepsilon) \times (A \cup \varepsilon) \times Q \times (P \cup \varepsilon)$ est une relation, appellée **relation de transition**

Remarque 6 : Une transition $t \in R$ est donc un quintuplet $(q_1, \gamma_1, a, q_2, \gamma_2)$ où
- $q_1, q_2 \in Q$ sont respectivement l'origine et l'extrêmité de la transition
- a est l'étiquette de la transition
- $\gamma_1 \in P$ est le symbole à dépiler
- $\gamma_2 \in P$ est le symbole à empiler

On peut représenter t de la façon suivante : $(q_1, \gamma_1) \xrightarrow{a} (q_2, \gamma_2)$ ou de façon graphique :



Définition 3 : Soit $\mathcal{A} = (A, Q, q_0, R, F, P)$ un automate à pile et $\alpha \in A^*$. On dit que α est reconnu par $\mathcal{A}$ ssi, la pile étant vide avant la lecture de α, il existe une suite non vide de transitions dont l'origine est q_0, d'étiquette α aboutissant soit à un état final, soit à la pile vide.

Question 10 : Construire un automate à pile reconnaissant le langage
- $L_1 = \{\alpha \in A^* \mid nbOcc(\alpha, a) = nbOcc(\alpha, b)\}$
- $L_3 = \{a^n b^n \mid n \in \mathbb{N}\}$.