Langages Formels

Licence Informatique – ENS Cachan

Partiel du 21 mars 2013

durée 3 heures

Document autorisé : polycopié du cours.
Toutes les réponses devront être correctement justifiées.
Le chiffre en regard d'une question est une indication sur sa difficulté ou sa longueur.

1 Automates et logique

On fixe un alphabet Σ contenant au moins les deux lettres a et b. On s'intéresse à la formule

$$\varphi_1(x) = \exists z\,(z \le x \wedge P_b(z) \wedge \forall y\,(z < y \le x \to P_a(y)))\ .$$

On note $\Sigma' = \Sigma \times \{0,1\}$, $\Sigma_0 = \Sigma \times \{0\}$ et $\Sigma_1 = \Sigma \times \{1\}$. On rappelle qu'un mot $W = (a_1,b_1)\cdots(a_n,b_n) \in \Sigma_0^*\Sigma_1\Sigma_0^*$ code le mot $w = a_1\cdots a_n \in \Sigma^*$ et la valuation σ vérifiant $\sigma(x) = i$ si et seulement si $b_i = 1$. On écrit $W = (w,\sigma)$ pour indiquer que W code le mot w et la valuation σ. Le langage défini par φ_1 est

$$L_1 = \{W = (w,\sigma) \in \Sigma_0^*\Sigma_1\Sigma_0^* \mid w,\sigma \models \varphi_1\}\ .$$

[1] **a)** Donner un automate qui reconnaît L_1.

[1] **b)** Donner une expression sans étoile pour le langage L_1.

Pour tout mot $a_1\cdots a_n \in \Sigma^+$ et toute position $1 \le i \le n$, on note

$$(a_1\cdots a_n, i) = (a_1,0)\cdots(a_{i-1},0)(a_i,1)(a_{i+1},0)\cdots(a_n,0) \in \Sigma_0^*\Sigma_1\Sigma_0^*\ .$$

Pour tout langage $L \subseteq \Sigma_0^*\Sigma_1\Sigma_0^*$, on définit le langage

$$L' = \{(a_1,b_1)\cdots(a_n,b_n) \in (\Sigma')^* \mid \forall\, 1 \le\ i \le n,\ b_i = 1 \iff (a_1\cdots a_n, i) \in L\}\ .$$

[4] **c)** Montrer que si L est reconnaissable alors L' est aussi reconnaissable.
Indication : Si $\mathcal{A} = (Q,\Sigma',\delta,q_0,F)$ est un automate déterministe complet reconnaissant L, construire un automate $\mathcal{A}' = (Q',\Sigma',\delta',q_0',F')$ avec $Q' = Q \times 2^Q \times 2^Q$, $q_0' = (q_0,\emptyset,\emptyset)$ et $F' = Q \times 2^F \times 2^{Q\setminus F}$.

2 Automates d'arbres

Soit $A_2 = \{d_1,d_2\}$ un alphabet de directions et Σ un alphabet. Soit $t : A_2^* \to \Sigma$ un arbre binaire. Un ensemble $X \subseteq \mathrm{dom}(t)$ est un *préfixe* de t si $X \neq \emptyset$ et X est fermé par préfixe : $uv \in X$ implique $u \in X$ pour tous $u,v \in A_2^*$.

Pour un arbre t, un préfixe X de t et un nœud $u \in X$, on définit la frontière $\mathrm{Fr}(u, X, t)$ inductivement par :
– si u est une feuille de X, i.e., $uA_2 \cap X = \emptyset$, alors $\mathrm{Fr}(u, X, t) = t(u) \in \Sigma$,
– si u est d'arité 1 dans X, i.e., $uA_2 \cap X = \{ud\}$, alors $\mathrm{Fr}(u, X, t) = \mathrm{Fr}(ud, X, t)$,
– si u est d'arité 2 dans X, i.e., $uA_2 \cap X = \{ud_1, ud_2\}$, alors $\mathrm{Fr}(u, X, t) = \mathrm{Fr}(ud_1, X, t)\mathrm{Fr}(ud_2, X, t)$.

Soit $L \subseteq \Sigma^*$ un langage de mots. On note $L' \subseteq T_2(\Sigma)$ l'ensemble des arbres $t \colon A_2^* \to \Sigma$ tels qu'il existe un préfixe X de t avec $\mathrm{Fr}(\varepsilon, X, t) \in L$.

[4] **a)** Montrer que si $L \subseteq \Sigma^*$ est un langage reconnaissable de mots, alors $L' \subseteq T_2(\Sigma)$ est un langage d'arbres reconnaissable.

3 Grammaires

Pour cet exercice, on rappelle que le langage de Dyck

$$D_1^* = \{v \in \{a, b\}^* \mid |v|_a = |v|_b \text{ et } |u|_a \geq |u|_b \text{ pour tous préfixes } u \text{ de } v\}$$

est engendré par la grammaire non ambiguë $S \to aSbS + \varepsilon$.

[2] **a)** Montrer que le langage $L_1 = \{v \in \{a, b\}^* \mid |u|_a = |u|_b\}$ n'est pas rationnel.
Donner une grammaire algébrique G_1 qui engendre le langage L_1.
Il faut bien sûr prouver que G_1 engendre exactement le langage L_1.

[2] **b)** On considère le langage $L_1' = \{v \in L_1 \mid \forall u \in L_1,\ u \leq v \implies (u = \varepsilon \lor u = v)\}$.
Donner une grammaire algébrique *non ambiguë* G_1' qui engendre le langage L_1'.
Il faut bien sûr prouver que G_1' est non ambiguë et engendre L_1'.

[1] **c)** Montrer que le langage $L_2 = \{a^i b^j c^k \mid k = \max(i, j)\}$ n'est pas algébrique.

[2] **d)** Soit $L_3 = \{v \in \{a, b, c\}^* \mid |v|_a = |v|_b = |v|_c > 0\}$.
Donner une grammaire contextuelle G_3 qui engendre L_3.
Il faut bien sûr prouver que G_3 engendre exactement le langage L_3.

On note $\mathrm{pref}(L)$ l'ensemble des préfixes des mots d'un langage L.

[2] **e)** Montrer que si L est un langage algébrique alors $\mathrm{pref}(L)$ est aussi algébrique.

Soit $L \subseteq \Sigma^*$ un langage sur l'alphabet Σ. On note $\frac{1}{2}(L)$ l'ensemble des mots $u \in \Sigma^*$ tels que $u\Sigma^{|u|} \cap L \neq \emptyset$, i.e., l'ensemble des premières moitiés des mots de L.

[1] **f)** Montrer que $\frac{1}{2}(D_1^*) = \mathrm{pref}(D_1^*)$.

[4] **g)** Montrer que la moitié d'un langage rationnel est un langage rationnel : si $L \subseteq \Sigma^*$ est un langage rationnel alors $\frac{1}{2}(L)$ est aussi rationnel.

[**] **h)** Question subsidiaire : Montrer que la moitié d'un langage algébrique n'est pas nécessairement algébrique.