TD sur la théorie des langages 12 octobre 2007

Prérequis : Définition d'un alphabet, d'une lettre, d'un mot, d'un langage, de la loi produit sur les mots, d'un facteur, d'un préfixe, d'un suffixe, des opérations sur les langages et de leurs propriétés, d'un langage rationnel, d'une expression rationnelle, d'un code et des propriétés caractérisant les codes.
Durée : 1 h 50

TD 4 – Introduction à la théorie des langages et des codes

☞ Exercice 1

Soient A un alphabet et L, L_1, L_2, L_3 et L_4 cinq langages sur A. Montrer que :

(a) si $L_1 \subset L_2$ et $L_3 \subset L_4$ alors $L_1.L_3 \subset L_2.L_4$
(b) si $L_1 \subset L_2$ alors $L_1^* \subset L_2^*$
(c) $(L^*)^* = L^*$
(d) $L^+ = L.L^* = L^*.L$
(e) $L.L^* + \varepsilon = L^*$

☞ Exercice 2

Soient A un alphabet et L, L_1, L_2, L_3 quatre langages sur A. Parmis les propriétés suivantes, lesquelles sont vraies ?

(a) $L + \emptyset = \emptyset + L = L$
(b) $L_1 + L_2 = L_2 + L_1$
(c) $(L_1 + L_2) + L_3 = L_1 + (L_2 + L_3)$
(d) $L_1.L_2 = L_2.L_1$
(e) $L + A^* = A^* + L = A^*$
(f) $(L_1.L_2).L_3 = L_1.(L_2.L_3)$
(g) $\{\varepsilon\}.L = L.\{\varepsilon\} = L$
(h) $\emptyset.L = L.\emptyset = \emptyset$
(i) $L_1.(L_2 + L_3) = L_1.L_2 + L_1.L_3$

☞ Exercice 3

Soit $A = \{a, b\}$ un alphabet. Ecrire les expressions rationnelles (régulières) qui dénotent les langages suivants :

- L_1 = l'ensemble des mots de A^* se terminant par a
- L_2 = l'ensemble des mots de A^* contenant au moins un a
- L_3 = l'ensemble des mots de A^* contenant au plus un b
- L_4 = l'ensemble des mots de A^* contenant un nombre pair de a
- L_5 = l'ensemble des mots de A^* contenant autant de a que de b

Avez-vous rencontré une difficulté pour décrire l'un de ces langages ? Qu'en conclure ?

☞ Exercice 4

Soit $A = \{a, b\}$ un alphabet. Décrire les langages dénotés par les expression rationnelles suivantes :

- $e_1 = (a + b)^*$
- $e_2 = (a^*b^*)^*$
- $e_3 = a^*\, b\, a^*\, b\, a^*\, b\, a^*$

☞ Exercice 5

On rappelle qu'un langage non vide $C \subset A^+$ est un code si tout mot de C^* se décompose de manière unique comme produit de mots de C ainsi que les propriétés suivantes :

1. Tout sous-ensemble d'un code est un code.
2. Si C est un code, $\varepsilon \notin C$
3. Si C est un code, C et $\bigcup_{n \geq 2} C^n$ sont deux ensembles disjoints
4. Si tous les mots de C sont de même longueur non nulle, alors C est un code. On parle alors de code uniforme.
5. Si aucun mot de C n'est préfixe (resp. suffixe) d'un autre mot de C, C est un code. Dans ce cas, le code C est qualifié de code préfixe (resp. suffixe).

Lorsque ces propriétés ne sont pas suffisantes pour monter qu'un langage est ou n'est pas un code, on a recours à l'algorithme de Sardinas et Patterson qui consiste en la construction d'une suite d'ensembles :

- *Initialisation* : $U_0 = L^{-1}.L \setminus \{\varepsilon\}$
- *Itération* : $U_{n+1} = U_n^{-1}.L \cup L^{-1}.U_n$
- *Condition d'arrêt* : $\begin{cases} \varepsilon \in U_n & \Rightarrow L \text{ n'est pas un code} \\ U_n = U_{n-1} & \Rightarrow L \text{ est un code} \end{cases}$

Parmi les langages suivants, déterminer ceux qui sont des codes[1]

- $L_1 = \{a, b\}$
- $L_2 = ab^*$
- $L_3 = \{a, ab, ba\}$
- $L_4 = \{a, ba, bb\}$
- $L_5 = \{aa, baa, ba\}$
- $L_6 = \{a, abbba, babab, bb\}$
- $L_7 = a^+b^+$
- $L_8 = a^+b^*$
- $L_9 = \{a^n b^n, n \in \mathbb{N}^*\}$
- $L_{10} = (ab)^*$

☞ Exercice 6

Soit $A = \{0, 1\}$ un alphabet et C un sous-ensemble fini de A^n.

(a) Soient $\alpha, \beta \in A^n$, on pose $d_H(\alpha, \beta) =$ le nombre de rang entre 1 et n où les deux mots α et β diffèrent. Donner une définition plus formelle de cette quantité appelée distance de Hamming de α et β. Vérifier que l'on a bien les propriétés d'une distance.

(b) On pose $H(C) = \inf \{d_H(\alpha, \beta),\ (\alpha, \beta) \in C \times C,\ \alpha \neq \beta\}$. Cette quantité s'appelle constante de Hamming de C. Quelle utilité peut avoir cette constante ?

[1] On demande une démonstration pour chaque langage. Il est recommandé d'utiliser prioritairement les propriétés.