

Éléments de Théorie des Graphes

Graphes valués et automates

Éric SOPENA, eric.sopena@labri.fr

Sixième partie

Graphes valués et Automates

Alphabet, mot (1)

Un **alphabet** est un ensemble (fini) A de symboles appelés lettres.

Un **mot** m sur un alphabet A est une séquence (finie) de lettres prises dans A : $m = a_1 \dots a_k$.

On note A^* l'ensemble de tous les mots construits sur l'alphabet A .

Exemples.

$$A_1 = \{0, 1\}$$

$$m_1 = 00101011, \quad m_2 = 1101$$

$$A_2 = \{a, b, c\}$$

$$m_3 = baba, \quad m_4 = bac$$

Alphabet, mot (2)

Le nombre de lettres d'un mot m est sa **longueur**, notée $|m|$. Ainsi, pour $m = a_1 \dots a_k$, nous avons $|m| = k$. Il existe un unique mot de longueur nulle, le **mot vide**, noté ε .

On note A^n l'ensemble des mots de A^* de longueur n . En particulier, $A^0 = \{\varepsilon\}$ et $A^1 = A$.

Soit a une lettre de A et m un mot de A^* . Le **nombre d'occurrences** de la lettre a dans m , noté $|m|_a$, est le nombre de fois où la lettre a apparaît dans m .

$$|abaac|_a = 3, \quad |abaac|_d = 0, \quad |\varepsilon|_a = 0$$

Concaténation de mots

Soient u et v deux mots de A^* .

La **concaténation** de u et v est le mot, noté $u.v$ ou plus simplement uv , obtenu en « collant » le mot v à la suite du mot u .

Ainsi, $|uv| = |u| + |v|$ et, pour toute lettre a ,
 $|uv|_a = |u|_a + |v|_a$.
 $u = aba, v = cabac, uv = abacabac$

Pour tous mots u, v et w , nous avons :

- $\varepsilon u = u\varepsilon = u$ (ε est élément neutre)
- $u.vw = uv.w = uvw$ (associativité)

Langages

Un **langage** L sur un alphabet A est un sous-ensemble, fini ou infini, de A^* .

Par exemple, sur l'alphabet $A = \{a, b\}$, on peut définir les langages suivants :

- $L_1 = A^2 = \{aa, ab, ba, bb\}$,
- $L_2 = \{ \text{mots d'au plus quatre lettres ayant autant de } a \text{ que de } b \}$
 $= \{\varepsilon, ab, ba, aabb, abab, abba, baab, baba, bbaa\}$,
- $L_3 = \{ \text{mots ayant deux fois plus de } a \text{ que de } b \}$
(ce langage est infini).

Opérations sur les langages (1)

Soit A un alphabet. On définit les opérations suivantes sur les langages définis sur A^* :

Union.

$$L_1 \cup L_2 = \{ m \in A^* / (m \in L_1) \text{ ou } (m \in L_2) \}$$

Intersection.

$$L_1 \cap L_2 = \{ m \in A^* / (m \in L_1) \text{ et } (m \in L_2) \}$$

Produit (de concaténation).

$$L_1.L_2 = L_1L_2 = \{ m \in A^* / m = m_1m_2, m_1 \in L_1 \text{ et } m_2 \in L_2 \}$$

Opérations sur les langages (2)

Puissance.

$$L^0 = \{ \varepsilon \}$$

$$L^n = \{ m \in A^* / m = m_1m_2\dots m_n, \}$$

avec $m_i \in L$ pour tout $i, 1 \leq i \leq n$, pour $n \geq 1$

Étoile et "plus".

$$L^* = L^0 \cup L^1 \cup \dots \cup L^n \cup \dots$$

$$L^+ = L^1 \cup L^2 \cup \dots \cup L^n \cup \dots$$

(i.e. $L^+ = L.L^*$)

Langages rationnels (1)

Un langage sur un alphabet A est **rationnel** (on dit également **régulier**), s'il peut être construit à l'aide des opérations union, étoile et produit à partir des langages élémentaires composés du mot vide ou d'un mot à une seule lettre.

De façon inductive, un L.R. se définit donc ainsi :

- $\{\varepsilon\}$ est un L.R.,
- si $a \in A$, $\{a\}$ est un L.R.,
- si L est un L.R. alors L^* est un L.R. (L^+ et L^n aussi),
- si L_1 et L_2 sont des L.R., alors $L_1 \cup L_2$ et $L_1.L_2$ sont des L.R.

Langages rationnels (2)

Soit $A = \{a, b, c\}$. Les langages suivants sont rationnels :

- $L_1 = (\{\varepsilon\} \cup \{a\}).(\{b\} \cup \{c\})^*.\{a\}^3$
 $L_1 = \{aaa, bcbcaaaa, \dots, aaaa, abaaa, \dots\}$
- $L_2 = \{c\}^2.(\{a\} \cup \{b\})^* \cup (\{a\} \cup \{b\}).\{c\}^2)^3$
 $L_2 = \{ccaaa, cc, cca, ccaaaaa, cbcabcabc, \dots\}$
- $L_3 = (\{a\} \cup \{b\})^*.\{c\}$
 $L_3 = \{c, ac, bc, aac, abc, bac, bbc, aaac, \dots\}$

Expressions rationnelles

On utilise habituellement quelques conventions d'écriture qui permettent d'alléger les expressions précédentes :

- un langage singleton $\{a\}$ est noté simplement a ,
- l'union est notée simplement $+$.

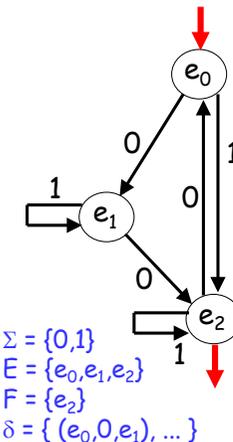
Ainsi, les langages précédents peuvent s'écrire, plus simplement, sous forme de ce que l'on appelle des **expressions rationnelles** :

- $L_1 = (\varepsilon + a)(b + c)^*a^3$
- $L_2 = c^2(a + b^* + (a + bc)^2)^3$
- $L_3 = (a + b)^*c$

Automates finis déterministes

Un **automate fini, déterministe** (ou A.F.D.) est un 5-uplet $\langle \Sigma, E, e_0, F, \delta \rangle$ où :

- Σ est l'**alphabet d'entrée**,
- E est un ensemble fini d'**états**,
- $e_0 \in E$ est l'**état initial**,
- $F \subseteq E$ est l'ensemble des **états terminaux**,
- $\delta : E \times \Sigma \rightarrow E$ est la **fonction de transition**.

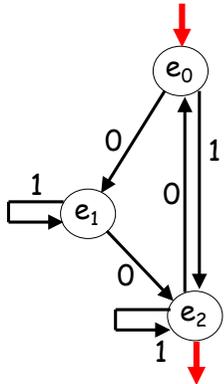


Langage reconnu par un A.F.D. (1)

On peut étendre la fonction δ aux mots en posant :

$$\delta(e, \varepsilon) = e$$

$$\delta(e, m) = \delta(\delta(e, a), m'), \text{ si } m = am'$$



Soit $m = 1101001$

$$\delta(e_0, m) = e_1$$

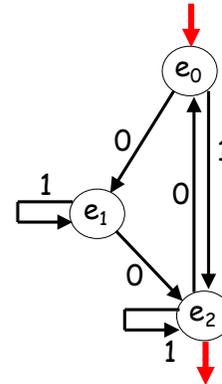
	1	1	0	1	0	0	1
e_0	e_2	e_2	e_0	e_2	e_0	e_1	e_1

Langage reconnu par un A.F.D. (2)

On peut étendre la fonction δ aux mots en posant :

$$\delta(e, \varepsilon) = e$$

$$\delta(e, m) = \delta(\delta(e, a), m'), \text{ si } m = am'$$



Le langage reconnu par un A.F.D. est alors :

$L = \{ m \in A^* / \delta(e_0, m) \in F \}$,
soit l'ensemble des mots de A^*
qui conduisent de l'état initial
à l'un des états terminaux.

Théorème de Kleene

Théorème [Kleene, 1956]. Un langage est rationnel si et seulement si il est reconnaissable.

Exemple. Le langage reconnu par notre automate est :

$$((1 + 01^*0)1^*0)^*(1 + 01^*0)1^*$$

