Théorie des langages et des automates

Ch1: Mots et langages

Olfa MOUELHI
olfa.mouelhi@esprit.tn
Ecole Supérieure Privée
d'Ingénierie et de Technologies

Année universitaire 2013/2014

DÉFINITIONS - ALPHABET

- Un alphabet Σ est un ensemble dont les éléments sont des symboles (lettres par exemple).
- Les alphabets sont toujours supposes finis.
- Exemples :
  - $\Sigma_1 = \{0; 1\}$
  - $\Sigma_2 = \{A; C; G; T\}$
  - $\Sigma_3 = \{a, b, c, \ldots, x, y, z\}$: l'ensemble de toutes les lettres (minuscules)
  - $\Sigma_4 = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, -, \times, /, (,)\}$.
  - $\Sigma_5 = \{\spadesuit, \clubsuit, \heartsuit, \diamondsuit\}$
  - etc.

DÉFINITIONS - MOTS

- Un mot (ou encore chaîne) w est une suite finie de symboles d'un même alphabet que l'on note par simple juxtaposition : $w = a_1a_2 \dots a_n;\ a_i \in \Sigma$
- Exemples :
  - $w_1 = \boldsymbol{10110}$ est un mot de $\boldsymbol{\Sigma_1}$
  - $w_2 = \boldsymbol{ordinateur}$ est un mot de $\boldsymbol{\Sigma_3}$
  - $w_3 = \boldsymbol{(7 - 6) \times 3}$ est un mot de $\boldsymbol{\Sigma_4}$

DÉFINITIONS - MOTS

- La longueur d'un mot w est le nombre de symboles qui le composent, et est notée $|w|$.
- Le mot vide, noté ε, est composé de 0 (zéro) symboles.
- $|\varepsilon| = 0$
- Le produit de concaténation de deux mots $x = a_1a_2 \ldots a_n$ et $y = b_1b_2 \ldots b_m$ est le mot xy obtenu par juxtaposition : $xy = a_1a_2 \ldots a_n b_1b_2 \ldots b_m$
- $|xy| = |x| + |y|$.

DÉFINITIONS - MOTS

- Σ^n est l'ensemble de toutes les chaînes de longueur n.
- $\Sigma^0 = \{\varepsilon\}$; cet ensemble n'est pas vide. Il contient la chaîne vide ε.
- Σ^* est la réunion des Σ^n pour $n \geq 0$. C'est donc l'ensemble de toutes les chaînes, chaîne vide comprise.
- La fermeture de Kleene d'un alphabet est l'ensemble de tous les mots de longueur quelconque de Σ, on la note Σ^*.
- Σ^+ est la réunion des Σ^n pour $n > 0$.
- L'ensemble des mots non vide construits sur Σ est la fermeture positive de Σ et est noté Σ^+.

Définitions - Mots

- Un mot x est une sous chaîne d'un mot w si et seulement si il existe deux mots y et z tels que $w = yxz$. Les mots w, x, y et z doivent appartenir à un même alphabet.
- Un mot x est préfixe d'un mot w si et seulement si il existe un mot y définit sur le même alphabet Σ que x et w tel que $w = xy$.
- Un mot x est suffixe d'un mot w si et seulement si il existe un mot y définit sur le même alphabet Σ que x et w tel que $w = yx$.

DÉFINITIONS – MOTS

Concaténation : pour tout $m, n \in \Sigma^*$, l'opération de concaténation $\bullet$ est définie par :

- $\bullet : V^m \times V^n \to V^{m+n}$

$$(x_1 \ldots x_m, y_1 \ldots y_n) \to x_1 \ldots x_m \bullet y_1 \ldots y_n = x_1 \ldots x_m y_1 \ldots y_n$$

- La concaténation est associative : $(w_1 \bullet w_2) \bullet w_3 = w_1 \bullet (w_2 \bullet w_3)$
- ε est l'élément neutre pour la concaténation $w \bullet \varepsilon = \varepsilon \bullet w = w$

Exemple

$\Sigma = \{a, b\} \qquad w = babb$

- Les préfixes de $w = \varepsilon, b, ba, bab, babb$
- Les suffixes de $w = \varepsilon, b, bb, abb, babb$

DÉFINITIONS – MOTS

Puissance : soit un alphabet Σ et $w \in \Sigma^*$,

$$w^n = \begin{cases} \varepsilon \text{ si } n = 0 \\ w \bullet w^{n-1} \end{cases}$$

Exemple

$\Sigma = \{a, b\}$ $\quad w = aab$

- $w^0 = \varepsilon$
- $w^1 = aab$
- $w^2 = aabaab$
- $w^3 = aabaabaab$

DÉFINITIONS – MOTS

- Soit Σ un alphabet. Soit A une partie de Σ. Pour tout mot $w \in \Sigma^*$, la longueur en A de w est le nombre d'occurrences de lettres de A dans le mot w. Ce nombre est note $|w|_A$.
- $|w| = |w|_\Sigma$.
- Pour tout symbole $\alpha \in \Sigma$, $|w|_\alpha$ le nombre d'occurrences du symbole α dans w.

Définitions - Mots

- Occurrence de symboles : nombre de fois où un symbole apparait dans un mot. On note $|w|_\alpha$ le nombre d'occurrences du symbole α dans w.
- Miroir : Soit $w = a_1 \dots a_n$, avec $a_1; \dots ; a_n \in \Sigma$. Le mot miroir de w est le mot note $\tilde{w}$ ou $w^{\sim}$ ou encore w^r défini par :

$$w = a_n \dots a_1$$

- $(uv)^{\sim} = \tilde{v}\tilde{u}$
- $(w^{\sim})^{\sim} = w.$

DÉFINITIONS - LANGAGE

- Un langage défini sur un alphabet Σ est une partie de Σ^* c'est donc un ensemble de mots défini sur Σ.
- Un langage L sur un alphabet Σ est un sous ensemble de Σ^* : $L \subset \Sigma^*$
- $L \in \mathcal{P}(\Sigma^*)$
- Les sous-ensembles de Σ^* sont appelés des langages formels.

  Exemple $\Sigma = \{a, b\}$, $\{a_n b_n / n \geq 0\}$ est un langage.
- Exemples triviaux :
  - $\varnothing$, le langage vide.
  - $\{\varepsilon\}$, le langage réduit à l'unique chaîne vide.
  - $L_1 = \{w \in \Sigma^* / w = w^{\sim}\}$
  - $L_2 = \{w \in \Sigma^* / |w| = 2k,\ k \geq 0\}$
  - $L_3 = \{w \in \Sigma^* / |w|_a = 2k,\ k \geq 0\}$

Opérations sur les langages

- Union

$$L_1 \cup L_2 = \{w \in \Sigma^* / w \in L_1 \text{ ou } w \in L_2\}$$

- Intersection

$$L_1 \cap L_2 = \{w \in \Sigma^* / w \in L_1 \text{ et } w \in L_2\}$$

- Complèmentaire

$$L^c = \Sigma^* \setminus L = \{w \in \Sigma^* \text{ et } w \notin L\}$$

- Différence

$$L_1 \setminus L_2 \text{ (ou } L_1 - L_2) = \{w \in \Sigma^* / w \in L_1 \text{ et } w \notin L_2\}$$

- Concaténation

$$L_1.L_2 = \{w \in \Sigma^* / \exists\, u \in L_1 \text{ et } v \in L_2 / w = u.v\}$$

$$L\{\varepsilon\} = \{\varepsilon\}L = L$$

$$L\,(M \cup N) = L(M) \cup L(N)$$

$$L\,(M \cap N) \subset L(M) \cap L(N)$$

Opérations sur les langages

- Puissances

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

  $L^1 = L$

  $L^{n+1} = L^n L \ (n \geq 1)$

  Si Σ est un alphabet alors Σ^n est l'ensemble des mots de longueur n.

  Itération (étoile) $L^* = \bigcup_{n \geq 0} L^n = \{w_1 \ldots w_n \,/\, n \geq 0$ et $w_1, \ldots w_n \in L\}$

  Itération stricte (plus) $L^+ = \bigcup_{n > 0} L^n = \{w_1 \ldots w_n \,/\, n > 0$ et $w_1, \ldots w_n \in L\}$

Opérations sur les langages: exemples

- $\Sigma = \{a, b\}$
- $L_1 = \{a, b\}$
- $L_2 = \{aa, bb, ab, ba\}$
- $L_3 = \{a, ab, bb\}$

$L_1 \cup L_2 = \{a, b, aa, bb, ab, ba\}$

$L_1 \cap L_2 = \emptyset$

$L_1 \cap L_3 = \{a\}$

$L_1 \setminus L_2 = \{a, b\}$

$L_3 \setminus L_1 = \{ab, bb\}$

$L_1.L_2 = \{aaa, abb, aab, aba, baa, bbb, bab, bba\}$

Opérations sur les langages : propriétés

Soient A, B et C trois langages définis sur un alphabet Σ

- $A.(B.C) = (A.B).C$
- $A \cup (B \cup C) = (A \cup B) \cup C$
- $A.(B + C) = A.B + A.C$
- $A.\ \emptyset = \emptyset$
- $A + \emptyset = A$
- $A \subseteq B \Rightarrow A.C \subseteq B.C$
- $A \subseteq B \Rightarrow C.A \subseteq C.B$