Graphes et Langages

Chapitre 5
Langages

Leo Donati   Noëlle Stolfi

Université de Nice Sophia Antipolis
IUT Nice Côte d'Azur
DUT Informatique

2015-2016

Chapitre 5 : Langages

Alphabet

Définitions

- Un alphabet est un ensemble non vide Σ de symboles.
- Les éléments de Σ sont appelés les lettres de l'alphabet Σ.

Exemples d'alphabet

- Alphabet binaire : les lettres sont les bits : $\Sigma = \{0, 1\}$
- Alphabet d'une seule lettre ; par exemple $\Sigma = \{a\}$
- Alphabet latin : $\Sigma = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\}$

Mots

Définition

Soit Σ un alphabet :

- les séquences finies de lettres de Σ s'appellent les mots sur l'alphabet Σ.
- le mot vide est aussi un mot, qui est formé par 0 lettres. Il est noté ε
- la longueur d'un mot M est égale à son nombre de lettres ; on la note $|M|$. Donc $|\varepsilon| = 0$.

Monoïde libre

Définition

Si Σ est un alphabet, on note par Σ^* l'ensemble de tous les mots possibles que l'on peut construire avec les lettres de cet alphabet. Σ^* s'appelle le monoïde libre sur Σ.

Remarque

Σ^* est toujours un ensemble infini, et on a toujours $\varepsilon \in \Sigma^*$.

Exemple

- Si $\Sigma = \{a\}$, alors $\Sigma^* = \{\underbrace{aaa \cdots aa}, \ k \in \mathbb{N}\}$
- Si $\Sigma = \{0, 1\}$, alors Σ^* est l'ensemble des mots binaires.

Concaténation

Définition

Soient M et N deux mots sur le même alphabet Σ, on peut construire un mot en écrivant à la suite les lettres de M puis les lettres de N. Ce nouveau mot est la concaténation des mots M et N et est noté $M \cdot N$.

Exemple

Sur l'alphabet $\Sigma = \{0, 1\}$, si $M = 1111$ et $N = 00$ alors

- $M \cdot N = 111100$
- $N \cdot M = 001111$
- $N \cdot N = 0000$

Opération de concaténation

Propriétés

La concaténation est une opération binaire sur le monoïde libre :

$$\cdot : \Sigma^* \times \Sigma^* \longrightarrow \Sigma^*$$

qui est :

- non commutative,
- associative,
- a comme élément neutre ε.

De plus :

$$|M \cdot N| = |M| + |N|$$

Notation puissance

Notations

- Si $M \in \Sigma^*$, on note $M^2 = M \cdot M$.
- Si a est une lettre de Σ, on note $a^k = \underbrace{aa \cdots aa}$ et $a^0 = \varepsilon$.
- On notera donc Σ^2, l'ensemble de tous les mots de 2 lettres que l'on peut construire avec les lettres de l'alphabet Σ.

Exemple

Si $\Sigma = \{a, b\}$, alors

$$\Sigma^2 = \{aa, ab, ba, bb\}$$

Langages

Définition

Un langage L sur un alphabet Σ, est un ensemble de mots, donc un sous–ensemble du monoïde libre :

$$L \subset \Sigma^*$$

Remarques

- $\emptyset$ et Σ^* sont des langages.
- Les langages peuvent être finis ou infinis.

Exemples de langage

Exemples

- Sur l'alphabet $\Sigma = \{a\}$, soit le langage :

$$L = \{a^k, \quad k \text{ est pair}\}$$

- Sur l'alphabet $\Sigma = \{0, 1\}$, sont des langages :

$$\begin{array}{rcl} L' & = & \{ \text{ mots commençant par } 0\} \\ L'' & = & \{ \text{ mots contenant la chaîne } 00\} \\ \Sigma^4 & = & \{ \text{ mots de } 4 \text{ lettres } \} \end{array}$$

Chapitre 5 : Langages

Intersection de langages

Définition

Si L_1 et L_2 sont deux langages sur l'alphabet Σ, alors leur intersection est aussi un langage, noté

$$L_1 \cap L_2$$

qui contient tous les mots communs de L_1 et de L_2.

Propriétés de l'intersection

Propriétés

- l'intersection de langages est commutative et associative,
- $L \cap L = L$,
- $L \cap \emptyset = \emptyset$,
- $L \cap \Sigma^* = L$,
- $L_1 \cap L_2 = L_1$ si et seulement si $L_1 \subset L_2$.

Union de langages

Définition

Si L_1 et L_2 sont deux langages sur l'alphabet Σ, alors leur union est aussi un langage, noté

$$L_1 \cup L_2$$

qui contient tous les mots présents dans au moins l'un des deux langages L_1 ou L_2.

Notation

An théorie des langages on préfère noter cette opération $L_1 + L_2$ au lieu de l'union et on parlera alors (improprement) de somme de langages.

Propriété de la somme

Propriétés

- La somme de langages est commutative et associative,
- $L + L = L$,
- $L + \emptyset = L$,
- $L + \Sigma^* = \Sigma^*$,
- $L_1 + L_2 = L_1$ si et seulement si $L_2 \subset L_1$.

Langage complémentaire

Définition

Si L est un langage sur l'alphabet Σ, alors le complémentaire de L, noté $\overline{L}$, est le langage qui contient tous les mots qui ne sont pas dans L.

Différence de langages

Si L_1 et L_2 sont deux langages sur l'alphabet Σ, alors la différence $L_1 \setminus L_2$ est le langage qui contient les mots de L_1 qui ne sont pas des mots de L_2.

Propriétés du complémentaire

Propriétés

- $\overline{L} = \Sigma^* \setminus L$,
- $\overline{\emptyset} = \Sigma^*$,
- $\overline{\Sigma^*} = \emptyset$,
- $\overline{\overline{L}} = L$,
- $\overline{L_1 + L_2} = \overline{L_1} \cap \overline{L_2}$,
- $\overline{L_1 \cap L_2} = \overline{L_1} + \overline{L_2}$,

Produit de langages

Définition

Soient L_1 et L_2 deux langages sur le même alphabet Σ ; le produit de deux langages, noté $L_1 \cdot L_2$, est le langage composé par *tous* les mots obtenus en concaténant un mot de L_1 et un mot de L_2 :

$$L_1 \cdot L_2 = \{M_1 \cdot M_2, \forall M_1 \in L_1, \forall M_2 \in L_2\}$$

Propriétés du produit

Propriétés

- le produit de langages n'est pas commutatif ;
- le produit de langages est associatif,
- le produit se distribue sur la somme de langages
- $L \cdot \emptyset = \emptyset$,
- $L \cdot \{\varepsilon\} = L$.

Opération étoile

Puissances

Soit L un langage sur l'alphabet Σ, on définit :

- $L^0 = \{\varepsilon\}$,
- $L^2 = L \cdot L$,
- $L^k = L^{k-1} \cdot L$

Étoile de Kleene

On définit le langage L^* comme la somme de toutes les puissances de L :

$$L^* = L^0 + L + L^2 + \cdots L^k + \cdots = \sum_{k=0}^{\infty} L^k$$

Opération plus

Définition

Alors que dans L^* on met aussi L^0, on peut aussi définir L^+ dans lequel on ajoute toutes les puissances de L à partir de 1 :

$$L^+ = L + L^2 + \cdots L^k + \cdots = \sum_{k=1}^{\infty} L^k$$

Lien entre L^+ et L^*

- $L^+ = L \cdot L^*$
- $L^* = L^+ + \{\varepsilon\}$
- Si L contient le mot vide, $L^+ = L^*$.

Propriétés de l'étoile

Propriétés

- $\{\varepsilon\}^* = \{\varepsilon\}$,
- $\emptyset^* = \{\varepsilon\}$
- $(L^*)^* = L^*$
- $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$ c'est l'ensemble de tous les mots non vides sur l'alphabet Σ.

Chapitre 5 : Langages

Expression régulières

Principe

Les expressions régulières sont des formules pour décrire, en peu de termes, la forme des mots qui composent un langage.

Exemple

Sur l'alphabet $\Sigma = \{0, 1\}$ on considère :

- $0(1^*0^*)$ est une expression régulière qui décrit le langage formé des mots binaires
  - commençants par un 0,
  - suivi d'un nombre quelconque de 1 (y compris aucun),
  - suivis d'un nombre quelconque de 0 (y compris aucun).
- $(0+1)^*$ décrit Σ^*,
- $(1+10)^*$: ?

Règles de formation

Règles de base

- $\emptyset$ est une expression régulière qui décrit le langage vide
- ε est une expression régulière qui décrit le langage $\{\varepsilon\}$,
- si $a \in \Sigma$, a est une expression régulière qui décrit le langage $\{a\}$,

Règles de composition

- Si E est une expression régulière qui décrit L , alors $(E)^*$ est une expression régulière qui décrit le langage L^*
- Si E et F sont des expressions régulières qui décrivent L_1 et L_2 , alors $(E) + (F)$ est une expression régulière qui décrit $L_1 + L_2$
- Si E et F sont des expressions régulières qui décrivent L_1 et L_2, alors $(E) \cdot (F)$ est une expression régulière qui décrit $L_1 \cdot L_2$

Langages réguliers

Proposition

Toute expression régulière décrit un langage.

Définition

Un langage régulier est un langage qui peut être décrit par une expression régulière.

Remarque

Les opérations d'intersection et de complémentaire ne sont pas autorisées dans les expressions régulières.

Propriétés des langages réguliers

Propriétés

Grâce aux règles de formation on peut déduire que :

- Le langage formé par un seul mot est régulier
- L'union, le produit et l'étoile d'un langage régulier est encore un langage régulier
- Tous les langages finis sont réguliers.

En revanche, il ne ressort pas automatiquement :

- ni que l'intersection de langages réguliers est régulier
- ni que le complémentaire d'un langage régulier est régulier.

Caractérisation des langages réguliers

Théorème

La classe des langages réguliers est la plus petite classe de langage qui contient les langages finis et qui est close pour les opérations suivantes :

- l'union (la somme),
- le produit,
- l'opération étoile.

Questions

Remarques

Le théorème précédent nous donne une idée de ce que les expressions régulières peuvent faire.
Mais il reste des questions qui ne sont pas résolues :

- l'intersection de langages réguliers est toujours régulier ?
- le complémentaire d'un langage régulier est toujours régulier ?
- Existe-t-il des langages non réguliers ? c'est-à-dire qui ne peuvent pas être définis par une expression régulière ?

Décidabilité

Étant donné une expression régulière, existe-t-il un procédé automatique qui soit capable de dire si un mot donné appartient ou pas au langage défini par l'expression ?