Alphabets, Mots, et langages

Chapitre 01

Alphabet

- Un alphabet est un ensemble fini non vide dont les éléments sont appelés lettres ou symboles.
- Un alphabet sera par exemple noté X ou Σ.

- Par exemple :

• $\Sigma = \{0,1\}$: alphabet des nombres binaires.

• $\Sigma = \{a,b,\ldots,z\}$: l'ensemble des lettres minuscules.

Mot

- Un mot (fini) ω sur l'alphabet Σ est une suite (finie) de lettres et est noté par simple juxtaposition :

$\omega = x1x2\ldots xn$ où $\forall i \in \{1,\ldots,n\}$, $xi \in \Sigma$. Par exemple :

- abbac et bccca sont deux mots sur l'alphabet $\{a,b,c\}$.
- 01101 est un mot sur l'alphabet $\{0,1\}$.

Longueur d'un mot

- La longueur d'un mot ω est le nombre de symboles constituant ce mot ; on la note $|\omega|$.

Ainsi,

- $|abbac| = 5$ et $|ba| = 2$.
- On définit aussi le nombre d'occurrence d'une lettre d de Σ dans ω; $|\omega|_d$
- Exemple :

• Soit $\omega = |00011001|$ un mot dans $\Sigma = \{0,1\}$ donc :
$|\omega|0=5$ et $|\omega|1=3$.

Le mot vide

- Le mot vide est un mot sans symbole et donc de longueur 0. Ce mot est représenté par le symbole ε ($|\varepsilon| = 0$).

Concaténation des mots

- Soient X un alphabet, $x \in \Sigma *$ est un mot de longueur m et y $\in \Sigma *$ un mot de longueur n, la concaténation de x et y notée xy est le mot de longueur n+m dont les m premiers symboles représentent un mot égal à x et les n derniers représentent un mot égal à y.

- Plus précisément si $x=a_1a_2\ldots a_m$ et $y=b_1b_2\ldots b_n$ alors

$xy= a_1a_2\ldots a_mb_1b_2\ldots b_n$.

Concaténation des mots

- Exemple :

Soit $x=01101$ et $y=001$, alors

$xy=01101001$ et $yx=00101101$.

- La concaténation est associative ($(xy)z=x(yz)$) mais généralement non commutative.
- Le mot vide est l'élément neutre pour la concaténation : $\varepsilon x = x\varepsilon = x$.

Puissances d'un alphabet :

- Soit Σ un alphabet, on note par Σ^k l'ensemble de tous les mots ayant une longueur donnée k sur cet alphabet.

- Exemples :

• $\Sigma^0 = \{\varepsilon\}$ quelque soit l'alphabet Σ.

• Si $\Sigma = \{a,b\}$ alors $\Sigma^1 = \{a,b\}$, $\Sigma^2 = \{aa,ab,ba,bb\}$,

$\Sigma^3 = \{aaa,aab,aba,abb,baa,bab,bba,bbb\}$, … etc.

Puissances d'un alphabet :

- L'ensemble de tous les mots sur Σ est noté $\Sigma *$.
- Par exemple $\{a, b, c\}* = \{\boldsymbol{\varepsilon}$, a, b, c, aa, ab, ac, ba, bb, bc, ca, cb, cc, aaa, aab,…}.

D'une autre façon :

- $\Sigma^* = \Sigma^{\mathbf{0}} \cup \Sigma^{\mathbf{1}} \cup \Sigma^{\mathbf{2}} \cup \ldots$
- Parfois on veut exclure le mot vide de l'ensemble des mots, l'ensemble des mots non vide sur l'alphabet X est noté X+.
- $\Sigma^+ = \Sigma^{\mathbf{1}} \cup \Sigma^{\mathbf{2}} \cup \Sigma^{\mathbf{3}} \cup \ldots$
- $\Sigma^* = \Sigma^+ \cup \{\boldsymbol{\varepsilon}\}$

Miroir d'un mot

- Soit Σ un alphabet et soit $\omega \in \Sigma^*$, $|\omega| = n$ et $\omega = \omega_0 \, \omega_1 \ldots\ldots\ldots \omega_n$, $n \geq 0$
- Le miroir de ω est $\omega^R = \omega_n \omega_{n-1} \ldots \omega_1 \omega_0$

- Palindrome :

• Soit Σ un alphabet et soit $\omega \in \Sigma^*$, on appelle palindrome tous les mots tels que : $\omega = \omega^R$

Exemple : $\omega = 0110 \rightarrow \omega^R = 0110 \rightarrow \omega = \omega^R$

Facteurs

- Soient f , g , h et u des mots de Σ *;
- g est facteur gauche ou préfixe de f s'il existe h tel que

f = gh.

- g est facteur gauche propre ou préfixe propre si h est différent du mot vide.
- h est facteur droit ou suffixe de f s'il existe g tel que f = gh.
- h est facteur droit propre ou suffixe propre si g est différent du mot vide.

Exemple:

- "bon" est un facteur gauche du mot "bonjour".
- "jour" est un facteur droit du mot "bonjour".
- "on" est un facteur propre du mot "bonjour

Sous-mot

- Un mot v est un sous-mot d'un mot u s'il est obtenu par effacement de certaines occurrences de lettres dans u.

Exemple :

- Les mots aaa et bb sont deux sous-mots de aabab sans être des facteurs.

Lemme de Levy

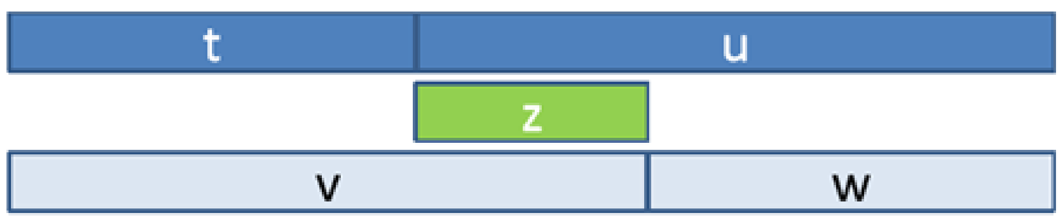
- Soient t, u, v et w quatre mots sur un alphabet X.

Si tu = vw alors il existe un unique mot $z \in X^*$ tel que :

- Soit t=vz et zu=w;
- Soit v=tz et zw=u.

Lemme de Levy : Preuve graphique

t=vz et zu=w



v=tz et zw=u

Langages

Langages

L'ensemble de tous les mots que l'on peut construire sur un alphabet Σ est noté Σ^*.

Un langage sur un alphabet Σ est un ensemble de mots construits sur Σ.

Tout langage défini sur Σ est donc une partie de Σ^*.

L'ensemble de tous les langages que l'on peut définir sur Σ^* est l'ensemble des parties de Σ^*, noté $\mathcal{P}(\Sigma^*)$.

Langages

Définition :

- On appelle langage (formel) tout sous-ensemble L de Σ^*, c'est-à-dire $L \subset \Sigma^*$.

Exemples :

- $L1 = \Sigma^*$, $L2 = \varnothing$, $L3 = \{\varepsilon\}$, $L4 = \{ \omega \in \Sigma^*, = \omega_1 ab \omega_2\}$
- Un langage peut être fini ou infini.

Soit $\Sigma = \{0,1\}$

- $L1 = \{\omega \in \Sigma^*, \omega 1 \equiv [3]\}$, L1 est infini.
- $L2 = \{\omega \in \Sigma^*, |\omega| < 5\}$, L2 est fini.

Langages

- Note :

Parmi les langages il convient de distinguer :

- L'ensemble $\emptyset$(l'ensemble vide qui ne contient aucun mot) ,
- Le langage $\{\varepsilon\}$ (le langage qui contient comme seul élément le mot vide).

Remarques :

- Un langage fini est un langage contenant un nombre fini de mots;
- Le langage vide ne contient aucun mot;
- Un langage est dit propre s'il ne contient pas le mot vide;
- Un langage est infini s'il n'est ni vide ni fini. Certains langages infinis (langages semi-décidables) peuvent être décrits par un ensemble de règles, appelé une grammaire formelle. Il existe d'autres langages infinis pour lesquels il n'existe aucun moyen de description, on les appelle des langages indécidables

Opérations sur les langage

- Soit X un alphabet, un certain nombre d'opérations peuvent être réalisées sur les langages :

- **L'union** : $\boldsymbol{L1} \cup \boldsymbol{L2} = \omega \in \boldsymbol{X*} \mid \omega \in \boldsymbol{L1}\ \boldsymbol{ou}\ \omega \in \boldsymbol{L2}$
- **L'intersection** : $\boldsymbol{L1} \cap \boldsymbol{L2} = \omega \in \boldsymbol{X*} \mid \omega \in \boldsymbol{L1}\ \boldsymbol{et}\ \omega \in \boldsymbol{L}$
- **Le complément** par rapport à X* : $\boldsymbol{L} = \omega \in \boldsymbol{X*} \mid \omega \notin \boldsymbol{L}$
- **La différence** : $\boldsymbol{L1} - \boldsymbol{L2} = \boldsymbol{L1} \cap \boldsymbol{L2} = \omega \mid \omega \in \boldsymbol{X*} \mid \omega \in \boldsymbol{L1}\ \boldsymbol{et}\ \omega \notin \boldsymbol{L2}$
- **La concaténation** : $\boldsymbol{L1}.\boldsymbol{L2} = \boldsymbol{u}.\boldsymbol{v} \mid \boldsymbol{u} \in \boldsymbol{L1}\ \boldsymbol{et}\ \boldsymbol{v} \in \boldsymbol{L2}$

Opérations sur les langage

- Puissance d'un langage : $\boldsymbol{L^2 = L.L \ et} \ \forall \boldsymbol{n} \in \mathbb{N}$,

$\boldsymbol{L^{n+1} = L^n. \ avec \ L^0 = \varepsilon}$

- (le passage à) l'étoile Kleene :

le langage L* est défini par :

– $\boldsymbol{L^* = L^0 \cup L \cup L^2 \ldots \cup L^n \ldots} =$

$$\{\boldsymbol{u} \mid \exists \boldsymbol{n} \in \mathbb{N}, \boldsymbol{u_1}, \ldots, \boldsymbol{u_n} \in \boldsymbol{L \ tel \ que \ u = u_1 \ldots u_n}\}$$

- L'opération plus : $\boldsymbol{L^+ = L^*.L = L \cup L^2 \ldots \cup L^n \ldots}$

Exemple

Soient L_1, L_2 et L_3 trois langages définis par :

$$L_1 = \{\varepsilon, aa\},\ L_2 = \{a^i b^j / i, j \geqslant 0\} \text{ et } L_3 = \{ab, b\}.$$

Calculer : $L_1.L_2, L_1.L_3, L_1 \cup L_2, L_2 \cap L_3, L_1^{10}, L_1^*, L_1^+, L_2^R$.

Solutions :

- $L_1.L_2 = L_2$;
- $L_1.L_3 = \{ab, b, aaab, aab\}$;
- $L_1 \cup L_2 = L_2$;
- $L_2 \cap L_3 = L_3$;
- $L_1^{10} = \{a^{2n} / 10 \geqslant n \geqslant 0\}$;
- $L_1^* = L_1^+ = \{a^{2n} n \geqslant 0\}$;
- $L_2^R = \{b^i a^j / i, j \geqslant 0\}$.

Propriétés

- La concaténation des langages n'est pas idempotante. cela signifie que $\exists L / L.L \neq L$;
- La concaténation des langages est associative ;
- La concaténation des langages n'est pas commutative ;
- La concaténation des langages est distributive par rapport à l'union ;
- La concaténation des langages n'est pas distributive par rapport à l'intersection ;
- $L^* = (L^*)^*$;
- $L^* = L^*.L^*$;
- $(L_1 + L_2)^* = (L_1^*.L_2^*)^* = (L_1^* + L_2^*)^*$.

Propriétés des opérations sur les langages

- Soit L, L1, L2, L3 quatre langages définis sur l'alphabet A :

  - $L^* = L^+ + \{\varepsilon\}$;
  - $L_1.(L_2.L_3) = (L_1.L_2).L_3$;
  - $L_1.(L_2 + L_3) = (L_1.L_2) + (L_1.L_3)$;
  - $L.L \neq L$;
  - $L_1.(L_2 \cap L_3) \neq (L_1 \cap L_2).(L_1 \cap L_3)$;
  - $L_1.L_2 \neq L_2.L_1$.
  - $(L^*)^* = L^*$;
  - $L^*.L^* = L^*$;
  - $L_1.(L_2.L_1)^* = (L_1.L_2)^*.L_1$;
  - $(L_1 + L_2)^* = (L_1^* L_2^*)^*$;
  - $L_1^* + L_2^* \neq (L_1 + L_2)^*$

Exemples de langages

$$\Sigma = \{a\} \qquad L_1 = \{\varepsilon, a, aa, aaa, \ldots\}$$

$$\Sigma = \{a, b\} \qquad L_2 = \{\varepsilon, ab, aabb, aaabbb, aaaabbbb, \ldots\}$$

$$\Sigma = \{a, b\} \qquad L_2 = \{\varepsilon, ab, aabb, aaabbb, aaaabbbb, \ldots\}$$

$$\Sigma = \{a, b\} \qquad L_3 = \{\varepsilon, aa, bb, aaaa, abba, baab, bbbb, \ldots\}$$

$$\Sigma = \{a, b, c\} \qquad L_4 = \{\varepsilon, abc, aabbcc, aaabbbccc, \ldots\}$$

Exemples de langages

- $\Sigma = \{0 \ldots 9, +, \times, /, -, (,)\}$
- Une expression arithmétique peut être représentée par un mot $m \in \Sigma^*$
- L est l'ensemble des expressions arithmétiques bien formées
  - $(1 + 3) \times 45 \in L$
  - $/234 \notin L$
  - $(1 + 4456 \notin L$

Exemples de langages

- $\Sigma = \{a, \ldots z, \vee, \wedge, \neg, \rightarrow, \leftrightarrow\}$
- Une formule de la logique des propositions peut être représentée par un mot $m \in \Sigma^*$
- L est l'ensemble des formules de la logique des propositions bien formées
  - $a \vee b \wedge c \in L$
  - $\vee b\ c \notin L$

Exemples de langages

- Σ est l'ensemble des identifiants, des constantes, des opérateurs et des mots clefs du langage C
- Un programme en langage C peut être représenté par un mot $m \in \Sigma^*$
- L est l'ensemble des programmes syntaxiquement corrects du langage C
  - `main(void){printf(‘‘hello world’’);}` $\in L$
  - `main(void){printf(‘‘hello world’’)}` $\notin L$

Description de langages

- **Description en langue naturelle par exemple :** .
- Langage L1 sur l'alphabet {0,1} : ensemble des mots dont l'interprétation comme entier est un multiple de trois. Il contient par exemple le mot : 1001 mais pas 1000.
- Langage L2 sur l'alphabet {a,b} : formé de tous les mots palindromes. Ainsi le langage L2 contient abbabba mais pas abbabab.
- Langage L1 sur l'alphabet {0,1} : ensemble des mots dont l'interprétation comme entier est un multiple de trois.

Description de langage

- **Les descriptions énumératives :**

- elles sont clairement utilisées pour les langages finis, mais également pour certains langages infinis tel que :
- L3 = $\{a^n b^n \ / \ n \geq 1\}$: contenant tous les mots formés exactement d'une suite de n occurrences de la lettre a suivi d'une suite contenant le même nombre d'occurrences n de la lettre b.

Description de langages

- **La définition par expression :**

• par exemple l'expression ab*cabc représente le langage L_4 dont les mots commencent par une occurrence de la lettre a, suivie par un nombre quelconque (éventuellement nul) d'occurrences de la lettre b, suivie du facteur droit cabc.

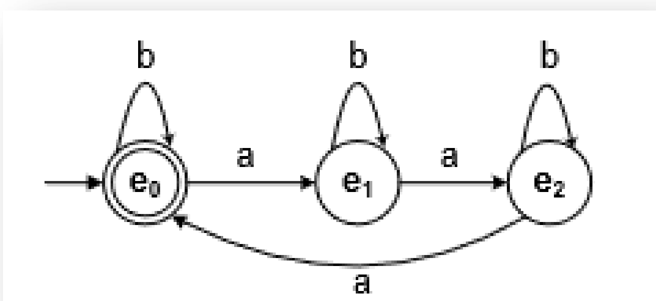
Description de langages

- **Les mécanismes génératifs :**
- appelés grammaire ou systèmes de réécriture : ils définissent un mécanisme de génération de mots sous la forme de règles de construction inductives.
- Le langage L_3 produit par les règles de la grammaire suivante: $S \longrightarrow aSb \;\mid\; S \longrightarrow ab$

Description de langage

Les mécanismes de reconnaissance :

- appelés aussi automates ou machine : ils permettent de dire si un mot appartient ou non au langage considéré.
- Exemple :
  - Automates d'états finis,
  - Automates à pile.



$\# \, S_0 \, a \rightarrow \# \, a \, S_0$

$\# \, S_0 \, b \rightarrow \# \, b \, S_0$

$a \, S_0 \, a \rightarrow a \, a \, S_0$

$a \, S_0 \, b \rightarrow S_0$

$b \, S_0 \, a \rightarrow S_0$

$b \, S_0 \, b \rightarrow b \, b \, S_0$

$\# \, S_0 \rightarrow \# \, S_f$

Merci pour votre attention