Chapitre 1

Mots et langages

DR ING FATMA SOMAA

CPI2

Plan

1. Vocabulaire et Mot
2. Langage
3. Propriétés des langages
4. Operations sur les langages
5. Définition des langages

Vocabulaire et mot

Un **alphabet** (ou vocabulaire) est un ensemble fini, non vide de symboles. On le note généralement **X ou** Σ

- Alphabet Latin $\Sigma=\{a,b,c,\ldots,z\}$
- Alphabet binaire $\Sigma=\{0,1\}$
- $\Sigma=\{rouge,noir,0,1,a\}$

Mot ou chaîne : Séquence de symboles de l'alphabet. Noté w.

- w_1= voiture ; w_2 = voyage deux mots définies sur l'alphabet Latin
- w_1=00101 ; w_2=101101 sont deux mots définies sur l'alphabet binaire.
- w_1=noir01rouge ; w_2=10AAnoir : sont deux mots définies sur l'alphabet $\Sigma=\{rouge,noir,0,1,a\}$

Vocabulaire et mot

Taille d'un mot : |w|= nombre de symboles constituant le mot.

- |rouge|=5 en considérant l'alphabet Latin
- |001|=3 en considérant l'alphabet binaire
- |rouge|=1 en considérant l'alphabet $\sum$={rouge,noir,0,1,a}

Chaîne vide : notée e s'il n'appartient pas à l'alphabet ou ε $|\varepsilon|=0$.

Sous chaîne : *x* est une sous chaîne de *w* si il existe *y* et *z* (chaînes sur la même alphabet). Tel que $w = y\,x\,z$.

Vocabulaire et mot

Préfixe : x est un préfixe de w si il existe y tel que: $w = x\, y$.

Suffixe : x est un suffixe de w si il existe y tel que $w = y\, x$.

- x = voit est un *préfixe* de w = voiture, car il existe y = ure tel que w = voit ure = $x\, y$
- x = ture est un *suffixe* de w = voiture, car il existe y = voi tel que w = voi ture = $x\, y$

Exercice

Ex1: Donner les sous chaines wi associes aux mots suivants :

1) abba sur le vocabulaire {a, b}

2) (x1*(x2+x1)) sur le vocabulaire {x1, x2, +, *, (,)}

Ex2:

Quelle est la longueur des mots abba et ε sur le Vocabulaire {a,b}

Exercice corrigé

<u>Ex1:</u>

1) Si le vocabulaire X = {a, b} alors dans le mot abba,

w_1 = a w_2 = ab w_3 = bb w_4 = abba …..

2) Si le vocabulaire est X = {x1, x2, +, *, (,)} alors dans le mot **(x1*(x2+x1))**

w_1 = (**w_2 = x1** **w_3 = *** **w_4 = (** **w_5 = x2**

w_6 = + **w_7 = x1** **w_8 =)** **w_9 =) ….**

Le mot abba est de longueur 4, |abba| = 4

Le mot e est de longueur 0, | ε | = 0

Vocabulaire et mot

Nombre d’occurrences d’un symbole dans un mot :

Le nombre d’occurrences d’un symbole x dans un mot w est le nombre de fois ou ce symbole apparait dans ce mot w. On le note $|w|_x$.

$$|\omega| = \Sigma_{x\in X} |\omega|_x$$

Exercice :

Quel est le nombre d’occurrences de b dans les mots abba et ε

Corrige :

$|abba|_b=2$

$|\varepsilon|_b=0$

Langage

Ensemble de mots choisis dans un alphabet.

Un langage peut être infini mais il existe un nombre finie de symboles permettant de composer les mots de ce langage.

$\sum^k$ = ensemble des mots de longueur k avec des symboles de $\sum$

Exemple : $\sum = \{0,1\}$

$\sum^1 = \{0,1\}$

$\sum^2 = \{00,01,10,11\}$

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

Langage

Σ^* Ensemble de toute les séquences de taille fini défini sur Σ : Fermeture de l'alphabet.

$$\Sigma^* = \{\varepsilon\} \cup \Sigma^1 \cup \Sigma^2 \ldots$$

$$\Sigma^+ = \Sigma^1 \cup \Sigma^2 \ldots$$

$$\Sigma^+ \cup \{\varepsilon\} = \Sigma^*$$

Un langage c'est un ensemble de mots appartenant à Σ^* et qui vérifie une propriété donnée :

L= $\{w \in \Sigma^* \mid w$ possède la propriété P$\}$

Langage

Exemples

- Ensemble des mots anglais légales

- Ensemble des mots de l'alphabet binaire contenant un nombre de *n* de 0 suivie par le même nombre *n* de 1.

  L={ε ; 01 ; 0011; 000111; … }.

- Ensemble des mots de l'alphabet binaire ayant un même nombre de 0 et de 1.

  L={ε ; 01 ; 10; 0101; 1001; … }.

- Ensemble des mots de l'alphabet binaire tel que leur valeur est premier.

  L={ε ; 10; 11; 101; 111; 1011; …}

Langage

Exemples

- Le langage vide L= $\varnothing$;
- Le langage $\{\varepsilon\}$ contenant le mot vide.
  - Note: $\varnothing \neq \{\varepsilon\}$.
  - Note: L'alphabet Σ est un ensemble fini.
- Ensemble des palindromes sur l'alphabet $\Sigma = \{a,b\}$
  - $L = \{w \in \Sigma^* \mid w = w^R\}$
  - $L = \{\varepsilon, aba, bab, a, b, \ldots\}$

Propriétés des langages

- Σ^* est infinie et dénombrable.
- $L = L_1 \cup L_2 = \{w \in \Sigma^* \mid w \in L_1 \text{ ou } w \in L_2\}$
- $L = L_1 \cap L_2 = \{w \in \Sigma^* \mid w \in L_1 \text{ et } w \in L_2 \}$
- Concaténation :

  $L = L_1 \,.\, L_2 = L_1L_2 = \{w \in \Sigma^* \mid \exists\, x , y , w = x\, y , x \in L_1, y \in L_2 \}$
- **Fermeture de Kleene.**
  - $L^* = \{w \in \Sigma^* \mid w = w_1w_2\ldots w_k,\ k \geq 0 \text{ et } w_1,w_2,\ldots,w_k \in L\}$
  - $k = 0 \Rightarrow w = \varepsilon$; $k = 1 \Rightarrow w \in L$;
  - Si L est un langage alors L* désigne l'ensemble de toute les chaînes de longueur finies formées par concaténation de mots de L, où chaque mot peut être utilisé de 0 à n fois, la chaîne vide est inclus.

Propriétés des langages

- $L \cup M = M \cup L$.

  Union est commutative.
- $(L \cup M) \cup N = L \cup (M \cup N)$.

  Union est associative.
- $(LM)N = L(MN)$.

  Concaténation est associative

  Note: Concaténation n'est pas commutative, i.e.,

  Il existe L et M tel que $LM \neq ML$.

Exemple :

- $L=\{aa,b\}$
- $L^*=\{\varepsilon,b,aa,bb,aab,baa,bbb,aaaa,aabb,baab,bbaa,bbbb,aaaab,aabaa,aabbb,baaaa,bbaab,bbbaa,bbbbb,\ldots\}$

Note : $\emptyset^* =\{\varepsilon\}\neq \emptyset$

Propriétés des langages

- $L(M \cup N) = LM \cup LN$.

  Concaténation est distributive à gauche pour l’union.

- $(M \cup N)L = ML \cup NL$.

  Concaténation est distributive à droite pour l’union.

- $L \cup L = L$.

  Union est idempotente.

- $\varnothing^* = \{\varepsilon\}$, $\{\varepsilon\}^*=\{\varepsilon\}$
- $L^+ = LL^* = L^*L$, $L^*= L^+ \cup\{\varepsilon\}$
- $(L^*)^*= L^*$. Fermeture est idempotente

Operations sur les langages

Concaténation : soient u et v deux mots définis sur l'alphabet Σ, tel que :

$$u = x_1 x_2 \ldots x_n \qquad v = y_1 y_2 \ldots y_m$$

Concaténation est non commutative

$$w = u.v = x_1 x_2 \ldots x_n y_1 y_2 \ldots y_m$$

Propriétés :

- $|w| = |uv| = |u| + |v|$
- *. Est associative*
- *ε est l'élément neutre pour la concaténation.* $\varepsilon x = x\varepsilon = x$

Operations sur les langages

***Facteur* :** soit u,v,w,t des mots définis sur $\sum$ tel que $w = uvt$

- **si** $u = \varepsilon$ alors v est dit facteur gauche de w (ou préfixe).
- **si** $t = \varepsilon$ alors v est dit facteur droit de w (ou suffixe).
- **si** $u = t = \varepsilon$ alors w est un facteur de lui même.

Occurrence d'un symbole dans un mot :

$$|abaaba|_a=4$$

Image (reverse) : $w=aabab$ $w^R=babaa$

Operations sur les langages

***Facteur* :** soit u,v,w,t des mots définis sur $\sum$ tel que $w = uvt$

- **si** $u = \varepsilon$ alors v est dit facteur gauche de w (ou préfixe).
- **si** $t = \varepsilon$ alors v est dit facteur droit de w (ou suffixe).
- **si** $u = t = \varepsilon$ alors w est un facteur de lui même.

Occurrence d'un symbole dans un mot :

$$|abaaba|_a=4$$

Image (reverse) : $w=aabab \quad w^R=babaa$

Operations sur les langages

Soient A et B deux langages alors on a les opérations suivantes:

Intersection $A \cap B = \{\omega /\ \omega \in A \text{ et } \omega \in B\}$

Union $A + B = \{\omega /\ \omega \in A \text{ ou } \omega \in B\}$

Complémentation $A - B = \{\omega /\ \omega \in A \text{ et } \omega \notin B\}$

Concaténation

$$A.B = \{\omega /\ \exists u \in A \text{ et } \exists v \in B \text{ et } \omega = u.v\}$$

Propriétés

Soient A, B, C des langages, on a

$$A.(B+C) = A.B + A.C$$

$$(A+B).C = A.C + B.C$$

$$A \subseteq B \quad \Rightarrow \quad AC \subseteq BC$$

$$CA \subseteq CB$$

Operations sur les langages

Notation

$a^k =$ aaaaaaaaa...... $a^2 = aa$

<------------------>

k fois

$a^0 = \varepsilon$

$a^* = \{a^i \,/\, i \geq 0\} = \{a^0, a^1, a^2, ...a^i, ...\}$

$a^+ = \{a^i \,/\, i > 0\} = \{a^1, a^2, ..., a^i, ...\}$

Operations sur les langages

Exercice

Calculer A* pour chacun des ensembles A suivants:

A = {a}

corrigé

Si $A = \{a\}$ alors $A^* = \{a\}^* = a^*$

Car $A^* = A^0 + A^1 + A^2 + \ldots A^i + \ldots$

$A^0 = \{\varepsilon\} = \{a^0\}$

$A^1 = AA^0 = \{a\}\{\varepsilon\} = \{a\} = \{a^1\}$

$A^2 = AA^1 = \{a\}\{a\} = \{aa\} = \{a^2\}$

$A^i = \{a^i\}$

$A^{i+1} = A\,A^i = \{a\}\{a^i\} = \{a^{i+1}\}$

....

$A^* = \{\varepsilon, a, aa, aaa, \ldots\} = \{a^0, a^1, a^2, a^3, \ldots\} = a^*$

Operations sur les langages

Exercice

Calculer $A.\phi$ et $A.\{\varepsilon\}$

Montrer qu'on n'a pas $A.(B\cap C) = A.B \cap A.C$

corrigé

$A.\phi = \phi.A = \phi$

$A.\{\varepsilon\} = \{\varepsilon\}.A = A$

On n'a pas $A.\cancel{(B\cap C) = A.B} \cap A.C$

Contre exemple : $A = \{\varepsilon, x\}$, $B = \{xy\}$, $C = \{y\}$

Car $A.(B\cap C) \neq A.B \cap A.C$

En effet $(B\cap C) = \phi$ donc $A. (B\cap C) = \phi$

Par contre $A.B = \{xy, xxy\}$ et $A.C = \{y, xy\}$ donc $A.B \cap A.C = \{xy\}$

Définition des langages

langages formels Tout sous ensemble de Σ^* dont les mots peuvent être définis de deux façons

Définition par propriété : Modélisation formelle d'une description naturel d'un langage.

Exemple :

L1 : {ensemble de mots définies sur {a,b} de longueur pair}

L1 = $\{w \in \{a,b\}^* \ / \ |w| = 2n \ ; \ n \geq 0\}$

Définition des langages

Définition récursive : Définition dans laquelle, un langage est définie sur lui même.

$L_2=\{w \in \sum *|\ w = a \text{ ou } w = aw_1;\ w_1 \in L_2\}=\{a,aa,\ldots,aaaa,\ldots\}$

$L_3=\{w \in \sum *|\ w = \varepsilon \text{ ou } w = w_1w_2;\ |w_1| =2 \text{ et } w_2 \in L_3\}$

$L_3 \equiv L_1$

Définition des langages

Expressions régulières :

$L_4=\{\varepsilon, x, xx, xxx, xxxx, \ldots\}$

soit $S = \{x\}$ alors $L_4=S^*$ ou $L_4=\{x\}^*$

Considérons l'étoile de la fermeture de Kleene appliquée à la lettre *x.*

x*

- x^* : indiquera une séquence quelconque de x qui peut être vide.
- $x^* = \varepsilon$ ou x ou xx ou $xxx\ldots$

L_4 = langage (x^*)

Définition des langages

- Considérons le langage

$$L = \{a, ab, abb, abbb, abbbb, \ldots\}$$

Toutes les chaînes constitués par un *a* suivi d'un nombre quelconque de *b*

On peut noter : L=Langage(*ab**)

Langage dans lequel les mots sont la concaténation d'un *a* (*a*) initial avec un nombre quelconque de *b* (*b**).

Appliquons l'étoile de Kleene à toute la chaîne *ab*, on aura : $(ab)^* = \varepsilon$ ou *ab* ou *abab ou ….*

Définition des langages

Le langage définit par l'expression :

*ab*a*

➔Ensemble de toutes les chaînes de *a* et de *b* qui ont au moins deux lettres, qui commencent et finissent par un *a.* et qui n'ont que des *b* ou rien à l'intérieur.

langage (*ab*a*)={*aa*,*aba*,*abba*,*abbba*,…}

Remarque :

- **Fausse description** : Ensemble de tous les mots qui commencent et puis finissent par *a* et qui n'ont que des *b* (ou rien) entre eux.
- Le mot *a* appartient a cette description.

Définition des langages

Le langage définit par l'expression :

$$a^*b^*$$

➔ Ensemble de toutes les chaînes de *a* et de *b dans lesquelles les a's viennent avant les b's.*

langage (a^*b^*)={ε ,*a*,*b*,*aa*,*bb*,*ab*,*bb*,*aaa*,*abb*,...}

Remarque :

- $a^*b^* \neq (ab)^*$
- Le langage à droite contient *abab* tandis que celui à gauche ne le contient pas

Définition des langages

si r_1 et r_2 sont deux expressions régulières alors ;

- $r_1 r_2$
- $r_1 \cup r_2$
- r_1^*

Sont des expressions régulières

Exemple 1 :

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

$L=\{w \in \Sigma^* \mid w$ contient la sous chaîne $aa\ \}$

$R = (a \cup b)^* aa\ (a \cup b)^*$

Exemple 2 :

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

$L=\{w \in \Sigma^* \mid w$ *ne* contient pas 3 b consécutifs $\}$

$R= (a \cup ba \cup bba)^*\ (\varepsilon \cup b \cup bb)$

Définition des langages

Théorème :

- Un langage L est dit régulier si et seulement si il existe une expression régulière qui le génère.

Propriétés Étant donné deux langages réguliers L_1 et L_2

- $L_1 \cup L_2$: est un langage régulier
- $L_1 . L_2$: est un langage régulier
- L_1* : est un langage régulier
- $L_1 \cap L_2 =$ est un langage régulier
- L^R : est un langage régulier
- $L_1 \backslash L_2$: est un langage régulier

Définition des langages

Deux expressions régulières α et β sont dites équivalentes si $L(\alpha) = L(\beta)$ Autrement s'ils génèrent le même langage

Exemple :

Langage de tous les mots qui ont au moins 2 a's peut être décrit par l'expression régulière :

$(a \cup b)^*a\,(a \cup b)^*a\,(a \cup b)^*$

Autre expression régulière

$b^*ab^*a(a \cup b)^*$

On peut noter :

$(a \cup b)^*a\,(a \cup b)^*a\,(a \cup b)^* = b^*ab^*a(a \cup b)^*$

Exercice

Donner les expressions régulières qui génèrent les langages suivants:

L1 = {ω∈{a,b}*, tel que ω contient exactement bbb}

L2 = {ω∈{a,b}*, tel que ω contient la sous chaîne bbb}

L3 = {ω∈{a,b}*, tel que ω contient seulement 3b, le reste c'est des a's}

L4 = {ω∈{a,b}*, tel que ω contient un nombre de a divisible par 3}

L5 = {ω∈{a,b}*, tel que ω contient un nombre paire de a}

L6= {{ω∈{a,b}*, tel que ω ne contient pas 3b Consécultifs}

L6 = {ω∈{a,b}*, tel que ω contient un nombre impaire de b}

L7 = {ω∈{a,b}*, tel que ω contient la sous chaîne aaa ou la sous chaîne bbb mais pas les deux en même temps}

Questions ?