<u>Chapitre1 :</u> Mots et langages

En théorie des langages, l'ensemble des entités élémentaires est appelé <u>*l'alphabet*</u>. Une combinaison d'entités élémentaires est appelé un <u>*mot*</u>. Un ensemble de mots forment un <u>*langage*</u>

I. Alphabets et Mots

1. Définitions

A. Un alphabet

Un alphabet noté A est un ensemble fini non vide de symboles.

<u>Exemples d'alphabets</u>

- A_1 = {a, b, c}
- A_2 = {1, 2, 3, 4}
- A_3= {IM, CM, ING}

B. Un mot

Un mot, défini sur l'alphabet *A,* est une suite finie d'éléments de *A.*

Exemple des mots

- Sur l'alphabet A_1 : aa, ab, ba, abc, ….
- Sur l'alphabet A_2 : 11, 12, 123, 1234, 4213, …
- Sur l'alphabet A_3 : IM, CM, IMCMING, ….

C. Longueur d'un mot

La longueur d'un mot *u* défini sur un alphabet A, notée |*u*|, est le nombre de symboles qui composent u.

Exemples
- |a| = 1,
- |123| = 3,
- |IMCMING|= 3

Le mot vide, noté ϵ , est défini sur tous les alphabets et il a une longueur nulle (0) (autrement dit, | ϵ | = 0).

- Définition de A^+ : on note A^+ l'ensemble des mots de longueur supérieure ou égale à 1 que l'on peut construire à partir de l'alphabet A.

A={x,y} ; A^+={x,y,xx,xy,yx,yy,xxx,xxy,xyx,xyy,yxx,yxy,yyx,yyy,…}

- Définition de $\underline{A^*}$: on note A^* l'ensemble des mots que l'on peut construire à partir de A, y compris le mot vide : $A^* = \{ \epsilon \} \cup A^+$
  A={x,y}
  $A^*=\{ \epsilon ,x,y,xx,xy,yx,yy,xxx,\dots\}$

2. Opérations sur les mots

Concaténation des mots :

Soient deux mots *u* et *v* définis sur un alphabet A. La concaténation de *u* avec *v*, notée *u:v* ou simplement *uv*, est le mot formé en faisant suivre les symboles de *u* par les symboles de *v*. On notera u^n le mot *u* concaténé *n* fois.

($u^0 = \epsilon$, $u^n = u:(u^{n-1})$ pour (n> 1).

$a^2=aa$

Par exemple

Sur l'alphabet A_1 si u = aabb et v = cc, alors uv = aabbcc et u^3 = aabbaabbaabb.

Préfixe, suffixe et facteur

Soient deux mots u et v définis sur l'alphabet A

- u est un **préfixe** de v si et seulement si, $\exists\, w \in A^*$ tel que v= uw.
- u est un **suffixe** de v si et seulement si, $\exists\, w \in A^*$ tel que v= wu.
- u est un **facteur** de v si et seulement si, $\exists\, w_1 \in A^*$ et $\exists\, w_2 \in A^*$ tel que $v= w_1uw_2$.

II. Langage

A. Définition

Un langage, défini sur un alphabet A, est un ensemble de mots définis sur A. Autrement dit, un langage est un sous-ensemble de A^*.
Deux langages sont particuliers et ils sont indépendants de l'alphabet A

- Le langage vide : $\{L = \emptyset \}$
- Le langage contenant le seul mot vide $L = \{ \epsilon \}$

Définition d'un langage par propriété mesurable

$L_1 = \{ w \in A^* / |w| = 2k, k \geq 0\}$ ➔ mots de long paire
A={x,y}
L1={ ϵ ,xx,xy,yx,yy, long4,long6,…}

$T_1 = \{ w \in A^* / |w| = 2k+1, k \geq 0\}$ ➔ mots de long impaire

$L_2 = \{ w \in \{a,b\}^* / d(w) = 0\}$

$d(w)= |w|_a-|w|_b$
A={a,b}
w=aa ; $|w|=2$ $|w|_a=2$ $|w|_b=0$

$\{a,b\}^*=\{ \epsilon, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, baa,$ ....
L2={ ϵ,ab,ba,aabb,abab,abba,bbaa,baba,baab,…}

Définition récursive d'un langage

La définition est dite récursive si elle fait référence au langage lui-même.
A={a,b}
$L = \{ w \in A^* / w=aw_1b; w_1\in L \text{ ou } w_1 = \epsilon \}$

Exemple:
$L = \{a^nb^n / n > 0\}$; $L' = \{ w \in A^* / w=aw_1b; w_1\in L \text{ ou } w_1 = \epsilon \}$
Montrer que $L = L' \Leftrightarrow L\subseteq L'$ et $L'\subseteq L$

$L\subseteq L' \Leftrightarrow \forall w (w\in L \Rightarrow w\in L')$

Rappel :
-Induction simple
-Induction généralisée
Pour prouver, on utilise la preuve par induction (récurrence).
On prouve par induction une proposition P sur un ensemble dénombrable E (Mq $\forall n\in N, P(n)$)

1) Preuve par induction simple (récurrence):
Soit la propriété P(n) : Si $w = a^n b^n$, $n > 0$ alors il existe $w1 \in L'$ tq $w=aw_1b$

- Base d'induction : Vérifions que P(1) est vraie
  w=ab =aw1b avec $w1= \epsilon$ ➔ $w\in L'$
- Etape d'induction :
  Montrons que $\forall n\in N, P(n)\Rightarrow P(n+1)$

  Supposons que P(n) est vraie et montrons que P(n+1) est vraie

  $W=a^{n+1}b^{n+1}=aa^nb^nb=aw1b$; $w1\in L'$

  $w=a^{n+1} b^{n+1}=a\, a^n b^nb=aw_1b$ avec $w1\in L'$ car P(n) est vraie
  ➔$w \in L'$

  Conclusion : $\forall n>0$; P(n) vraie

⇨ $L\subseteq L'$

2) Mq $L'\subseteq L$
Preuve par induction généralisée :
Soit la propriété P(l) : $\forall l\geq 0$(Si $w \in L'$avec $|w|=l$, alors $\exists$ n> 0 tel que $w = a^n b^n$)

- Base d'induction généralisée : Vérifions que $P(l_0)$ est vraie. l_0 est la plus petite longueur pour laquelle la propriété est vérifiée.

Vérifions P(2) :

W=ab ∈ L' ; ∃ n=1 tq w= $a^n b^n$

- ⇨ W ∈ L
- ⇨ P(2) est vraie

- Etape d'induction généralisée:
  Montrons que $\forall l(\forall l_0 \leq k < l ; P(k) \Rightarrow P(l))$

  (En d'autre termes, supposons que la propriété est vraie pour toutes les longueurs < l et montrons qu'elle est vraie pour l)

  W ∈ L' avec |w|=l

  W=aw1b avec w1∈ L'

  w1∈ L' ; |w1|=l-2<l

  ➔P(|w1|) est vraie ➔ w1= anbn

  |w1|=|w|-2 =l-2 <l

  ➔P(|w1|) est vraie par supposition

- ⇨ ∃ n tel que w1 = $a^n b^n$
  W=aw1b= a $a^n b^n$b= $a^{n+1} b^{n+1}$

  Conclusion : P(l) vraie

- ⇨ L'⊆L

Concl : L=L'

B. Opérations ensemblistes définies sur les langages

Soient deux langages L_1 et L_2 respectivement définis sur les alphabets A_1 et A_2.

- L'***Union*** de L_1 et L_2 est le langage défini sur $A_1 \cup A_2$ contenant tous les mots appartenant soit à L_1 ou L_2.

$$L_1 \cup L_2 = \{u \mid u \in L_1 \ ou \ u \in L_2\}$$

- L'***intersection*** de L_1 et L_2 est le langage défini sur $A_1 \cap A_2$ contenant tous les mots appartenant soit à L_1 et L_2.

$$L_1 \cap L_2 = \{u \mid u \in L_1 \ et \ u \in L_2\}$$

- Le **complément** du langage L défini sur A* est le langage C(L) contenant tous les mots de A* qui n'appartiennent pas à L

$$C(L)= \{u \in A^* | u \notin L \}$$

- La **différence** de L_1et L_2 est le langage contenant tous les mots appartenant à L_1 et ils n'appartiennent pas à L_2

$$L_1 - L_2 = \{u | u \in L_1 \; et \; u \notin L_2 \}$$

C. Produits de deux langages

Le produit ou la concaténation de deux langages L_1 et L_2 respectivement définis sur les alphabets A_1 et A_2 est le langage défini sur l'union des alphabets $A_1 \cup A_2$ contenant tous les mots formés par un mot de L_1 suivi d'un autre de L_2.

$$L_1.L_2 = \{ uv \mid u \in L_1 \; et \; v \in L_2 \}$$

Le produit de deux langages et associatif et non commutatif.
Exemple

Soient L1= {00, 11} et L2 = {0, 1, 01} définis sur A= {0,1}

L1.L2={000,001,0001,110,111,1101}

$L1^2$=L1.L1={00,11}{00,11}={0000,0011,1100,1111}

- **Puissance d'un langage**

Les puissances successives d'un langage L sont définies d'une manière récursive :

- $L^0 = \{ \epsilon \}$
- $L^n = L.L^{(n-1)} \;\; \forall \, n > 1$

Exemple

Si L = {00,11} alors L^2 = {0000, 0011, 1100, 1111}

- **Fermeture itérative d'un langage**

La fermeture itérative d'un langage L (appelée aussi fermeture de Kleene) est l'ensemble des mots formés par une concaténation des mots de L.

$L^* = \{u \mid \exists \, k \geq 0 \; et \; u_1, u_2, \dots, u_k \in L$ tel que $u = u_1 u_2 \dots u_k \}$
En d'autres termes :

$L^* = \bigcup_{i=0}^{i=k} L^i$

$L^+ = \bigcup_{i=1}^{i=k} L^i$
L={a}
L^0= { ϵ }
L^1={a}
L^2={aa}
…
L^*={ $\epsilon, a, aa, aaa, \ldots$ }
L^+={ $a, aa, aaa, \ldots$ }

<u>Exercice 1</u>

Soit l'alphabet A= {0, 1} on considère les deux langages L_1= {01^n | n ∈ N } et
L_2 = {0^n1 | n ∈ N }
Définir les langages L_1. L_2, $L_1 \cap L_2$ et L_1^2

<u>Correction</u>
L_1= {0, 01, 011, 0111,… }
L_2 = {1,01,001,0001,… }

- $L_1.L_2$ = {01^n0^m1| n ∈ N et m ∈ N }
- $L_1 \cap L_2$ = {01}

L_1= {0, 01, 011, 0111,… }
L_1= {0, 01, 011, 0111,… }
- $L_1^2 = L1.L1 = \{01^n01^m$|n ∈ N et m ∈ N }

III. Expressions régulières

Définition récursive

Soit A un alphabet
- ϵ et $\emptyset$ sont des expressions régulières
- Tout symbole a de A est une expression régulière
- Si r est une expression régulière alors (r) , r^+, r^* sont aussi des expressions régulières .
- Si r_1 er r_2 deux expressions régulières alors (r_1+r_2) noté aussi r_1|r_2 et r_1r_2 sont des expressions régulières.

Langage régulier

- Soit l'application L qui associe à une expression régulière un langage, définie de la manière suivante :

L : Reg(A) ➔ A^*

- $L(a) = \{a\}$ pour tout a de A, $L(\epsilon) = \{\epsilon\}$, $L(\emptyset) = \emptyset$
- $L(r_1|r_2) = L(r_1) \cup L(r_2)$, $L(r_1r_2) = L(r_1).L(r_2)$
- $L(r^*) = L(r)^*$ et $L(r^+) = L(r)^+$

Exemple

Soit A = {a, b, c }

- L(b) = {b}
- $L(a|c) = L(a) \cup L(c) = \{a, c\}$
- L(ac) = L(a) . L(c) = {ac}

- $L(c^*) = L(c)^* = \{c\}^* = \{\varepsilon, c, cc, ccc, cccc...\}$
- $L(c^+) = L(c)^+ = \{c\}^+ = \{c, cc, ccc, cccc...\}$
-
- $L(a|c^*) = L(a) \cup L(c^*) = \{a\} \cup \{c\}^* = \{\varepsilon, a, c, cc, ccc, cccc...\}$
- $L((a|b)^*) = (L(a) \cup L(b))^* = \{a, b\}^*$

  $= \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, baa, bab \ldots\}$
- $\{a, b\}^* = \{a, b\}^0 \cup \{a, b\}^1 \cup \{a, b\}^2 \cup \ldots$
- $\{a, b\}^{2} = \{a, b\} \cdot \{a, b\} = \{aa,ab,ba,bb\}$

- $L((ac)^+|b) = (L(a) . L(c))^+ \cup L(b) = \{ac\}^+ \cup \{b\} = \{b,ac,acac,acacac,\ldots\}$
- $L(a^+|(abc)^*|((b|a)c)^+) = L(a)^+ \cup L(abc))^* \cup L(\{b, a\}.c)^+$

  $= \{a, aa, aaa \ldots \varepsilon, abc, abcabc \ldots bc, ac, bcbc, acac\ldots\}$

Exercice

Alphabet : A={a,b,c}

1)E.R : les mots sur A contenant au moins 3 a.

➔ (a|b|c)*a(a|b|c)*a(a|b|c)*a(a|b|c)*

2)E.R : les mots sur A contenant exactement 3 a.

➔ (b|c)*a(b|c)*a(b|c)*a(b|c)*

3)E.R : les mots sur {a,b} ne contenant pas le facteur ab.

➔ { ε,a,b,aa,bb,ba,aaa,bbb,baa,bba,…}

⇨ a*|b*|b*a*

⇨ b*a*

4)les mots représentant les nombres binaires.

E.R : $(0|1)^+$

5) les nombres décimaux multiples de 5.

E.R : $(0|1|2|\ldots|9)^*(0|5)$