Université des Sciences et de la Technologie Houari Boumediene

Faculté d’Informatique

Cours Théorie des langages (THL)

Chapitre 2 : Les Notions Fondamentales de la Théorie des Langages

2ieme année ING 2023/2024

Dr H.BELHADI hib.belhadi@gmail.com

ALPHABET

Définition

Un alphabet X est **un ensemble fini et non vide**. Les éléments de cet ensemble sont appelés **des lettres** ou **symboles**.

Exemples :

- Alphabet binaire X={0, 1}
- Alphabet décimal X={0, 1,…., 9}
- Alphabet des gènes (ADN) X={A, T, C, G}
- Alphabet des expressions arithmétiques X={+, -, *, ÷, (,), 0, 1, …, 9} .

MOTS

Définition

Un mot sur un alphabet X est une suite finie, éventuellement vide, d'éléments de X.

Exemples :

Alphabet	Mots
{0, 1}	0, 10, 010001, 0011001, 111111
{A, C, G, T}	ATTGCT, TTTGTACGT, GTTTCA
{+, -, *, ÷, (,), 0, …, 9}	5+3, 7***, +***))), ((4+8) ÷2)

MOTS

Notations :

- Le mot vide (suite vide d'éléments) est noté ε .

- L'ensemble des mots formés à partir d'un alphabet X est noté **X***.
  Exemple Si X={a} alors X*={ε , a, aa, aaa, aaaa, ...}

- **X^+** est l'ensemble des **mots non vides**.
  On a $X^* = X^+ \cup \{\varepsilon\}$.

Remarque : Les ensembles X* et X+ sont infinis.

Operations sur les MOTS

1) Concaténation

Définition : Soient w_1 et w_2 deux mots de X*. On définit la concaténation comme la juxtaposition de w_1 et w_2 et on la note $w_1.w_2$ (ou w_1w_2). Ainsi, **si** $w_1 = a_1 \ldots a_n$ et $w_2 = (b_1 \ldots b_m)$

alors $w1.w2 = a_1 \ldots a_n \, b_1 \ldots b_m$

Remarques:

- $\varepsilon.w = w.\varepsilon = w$ (ε est l'élément neutre de la concaténation)
- La concaténation n'est pas commutative ($w1w2 \neq w2w1$)
- La concaténation est associative $(w1w2)w3 = w1(w2w3)$

Operations sur les MOTS

2) Longueur

Définition : On appelle longueur d'un mot w sur un alphabet X la somme des occurrences des différents symboles le constituant. Elle est notée **lg(w)** (ou **| w |**).

- **Formellement :** → $lg(\varepsilon)=0$

→ $lg(a)=1 \quad \forall a \in X$

→ $lg(a.w) = 1+lg(w), \ \forall a \in X, \ \forall w \in X^*$

Exemples : $lg(ab)=2$ $lg(aba)=3$ $lg(abb)=3$

Remarques: → $|w|_a$ désigne le nombre de a dans le mot w.

→ La fonction longueur est une application de X^* vers $\mathbb{N}$

Operations sur les MOTS

3) Miroir

Définition: On appelle mot miroir d'un mot w, noté Mir(w) ou w^R le mot obtenu en inversant les symboles de w.

Ainsi **si** $w=a_1 \ldots a_n$ **alors** $Mir(w)= a_n \ldots a_1$.

- **Formellement :** → $Mir(\varepsilon)= \varepsilon$

  → $Mir(a)=a \quad \forall a \in X$

  → $Mir(a.w) = Mir(w).a \quad \forall a \in X, \quad \forall w \in X^*$

Exemples : Le miroir du mot abbaa est aabba.

Le miroir du aba est le mot lui-même (mot palindrome)

Remarque : $(w^R)^R = w$

Operations sur les MOTS

4) Puissance

Définition: La puissance d'un mot est définie par récurrence de la manière suivante :

$w^0 = \varepsilon$

$w^{n+1} = w^n.w \quad \forall n \geq 1$

Exemple :

Les puissances du mot abb sont $\{\varepsilon$, abb, abbabb, abbabbabb, ...$\}$

Operations sur les MOTS

5) Factorisation

Définition: Soient v et w deux mots de X*.

- v est **facteur ou sous-mot** du mot w si et seulement s'il existe deux mots u_1, u_2 appartenant à X* tel que : **w**= u_1**.v.**u_2
- Le mot v est **facteur propre** du mot w si $\mathbf{u_1} \neq \varepsilon$ **et** $\mathbf{u_2} \neq \varepsilon$.
- Le mot v est **facteur gauche (ou préfixe)** de w si $\mathbf{u_1}=\varepsilon$.
- Le mot v est **facteur droit (ou suffixe)** de w si $\mathbf{u_2}=\varepsilon$.

Exemple : Soit le mot w= aabbba, nous avons :

Le mot v1= abb est facteur de w, c'est un facteur propre.

Le mot v2= aab est facteur gauche de w.

Le mot v3= ba est facteur droit de w.

LANGAGE

Définition: Soit X un alphabet. On appelle langage formel défini sur X tout sous-ensemble de X*.

Exemple :

- L1 = l'ensemble des mots de {a, b}* qui commencent par a

  = {a, aa, ab, aaa, aab, aba, abb,…..}

  = $\{aw \ / \ w \in \{a,b\}^*\}$

- L2 = l'ensemble des mots de {a, b}* de longueur strictement inférieure à 3

= {ε, a, b, aa, ab, ba, bb}

- L3 = l'ensemble des nombre décimaux divisibles par 3

= {0, 3, 6, 9, 12, 15, 18, 21, …}

LANGAGE

- Un langage **vide** est un langage qui ne contient aucun mot, il est noté $\emptyset$.
- Un langage **fini** est un langage qui **contient un nombre fini de mots**. Dans l'exemple précédent L2 est fini. Un langage fini peut être décrit par l'énumération des mots qui le composent.
- Un langage **infini** est un langage qui contient une infinité de mots. Dans l'exemple précédent L1 est infini. De façon générale, un langage infini peut être décrit par :
  1) application d'opérations à des langages plus simples,
  2) un ensemble de règles appelées grammaires.
- Un langage est dit **propre** **s'il ne contient pas le mot vide**.
- Le langage $\emptyset$ est différent du langage $\{\varepsilon\}$

LANGAGE

- 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.

Operations sur les LANGAGES

Les langages étant des ensembles, on peut effectuer sur eux les opérations définies sur les ensembles :

- Union : $L_1 \cup L_2 = \{w \,/\, w \in L_1$ **ou** $w \in L_2\}$
- Intersection : $L_1 \cap L_2 = \{w \,/\, w \in L_1$ **et** $w \in L_2\}$
- Complément : $\overline{L} = \{w \,/\, w \in X^* \text{ et } w \notin L\}$
- Différence : $L_1 - L_2 = \{w \,/\, w \in L_1$ **et** $w \notin L_2\}$
- Produit : $L_1 \times L_2 = \{(w_1, w_2) \,/\, w \in L_1$ **et** $w \in L_2\}$

Operations sur les LANGAGES

De plus, les opérations définies sur les mots peuvent être étendues aussi aux langages.

Soient deux langages L_1 et L_2 respectivement définis sur les alphabets X_1 et X_2 et soit L un langage défini sur l'alphabet X.

La concaténation de langages

$$L_1.L_2=\{w_1.w_2 \ / w_1 \in L_1 \text{ et } w_2 \in L_2\}$$

Remarques :

$$\emptyset.L_1=L_1.\emptyset = \emptyset$$

$$\{\varepsilon\}.L_1=L_1.\{\varepsilon\}=L_1$$

Operations sur les LANGAGES

- **Langage miroir** $L^R = \{w^R \,/\, w \in L\}$
- **Puissance concaténative** $L^o = \{\varepsilon\}$ et $L^{n+1} = L^n.L$
- **Fermeture itérative ou Etoile de Kleen** $L^* = L^o \cup L^1 \cup \ldots \cup L^k \cup \ldots = \cup_{i \geq o} L^i$
- **L'étoile propre (ou ε libre)** $L^+ = \cup_{i \geq 1} L^i$

Remarques : si $\varepsilon \in L$ alors $\varepsilon \in L^*$ et $\varepsilon \in L^+$
si $\varepsilon \notin L$ alors $\varepsilon \in L^*$ et $\varepsilon \notin L^+$