THÉORIES DES LANGAGES

Mr,HEMIOUD
hemourad@yahoo,fr
Université de Jijel
Département d'informatique

Langages réguliers

Langages réguliers

- les langages réguliers sont très utilisés en pratique.

- les algorithmes permettant de reconnaitre ces langages sont simples. « *Reconnaitre un langage L consiste à dire, pour un mot m donné, si m* $\in L$ *ou m* $\notin L$. »

Langages réguliers

- **Définition** . L'ensemble des langages réguliers (noté $\boldsymbol{L_R}$) est le plus petit ensemble vérifiant :
  - $\emptyset \in L_R$
  - $\varepsilon \in L_R$
  - $a \in L_R$ pour tout $a \in$ A
  - Si $L \subset L_R$ et $L' \subset L_R$ alors $L \mid L' \subset L_R$, $L.L' \subset L_R$ et $L^* \subset L_R$.
- ***NB*** . Les langages réguliers sont utilisés :
  - Dans la première phase de la compilation (analyse lexicale)
  - Dans la recherche de chaines de caractères dans des documents.

Expressions régulières

Les langages réguliers sont représentés, de manière précise et consiste, par ***les expressions régulières (ER)***

Définition. Soit un alphabet A, l'ensemble des expressions régulières sur A est défini comme suit :

- **Ø** est une expression régulière.
- ε est une expression régulière.
- ***a*** est une expression régulière pour tout $a \in A$
- Si α et β sont deux expressions régulières alors $\alpha|\beta$, $\alpha.\beta$ et α^* sont des expressions régulières.

Remarque :

- Les parenthèses ne sont pas indispensable si l'expression régulière souhaitée respecte les priorités des opérateurs.

- L'ensemble de toutes les expressions régulières sur un alphabet donné A constitue un langage sur l'alphabet :

$$ER_{\mathrm{A}} = \mathrm{A} \cup \{\, \emptyset \,,\, \varepsilon \,,\, . \,,\, + \,,\, * \,,\, (\,,\,) \,\}$$

Langages dénotés par des expressions régulières

Définition

Soit α et β deux expressions régulières, $L(\alpha)$ (resp. $L(\beta)$) désigne le langage dénoté par α (resp. β).

L'identification du langage à partir de l'expression régulière qui le dénote doit respecter les règles suivantes :

- $L(\emptyset) = \emptyset$,
- $L(\varepsilon) = \varepsilon$,
- $L(a) = a$ pour tout $a \in$ A,
- $L(\alpha \mid \beta) = L(\alpha) \mid L(\beta)$,
- $L(\alpha.\ \beta) = L(\alpha).L(\beta)$,
- $L(\alpha+) = L(\alpha)+$
- $L(\alpha*) = L(\alpha)*$

Lois algébriques sur les expressions régulières

- $(\varnothing + R) \equiv (R + \varnothing) \equiv R$ élément neutre de l'union
- $(\varepsilon . R) \equiv (R . \varepsilon) \equiv R$ élément neutre de la concaténation
- $(\varnothing . R) \equiv (R . \varnothing) \equiv \varnothing$ élément absorbant de la concaténation
- $R + S \equiv S + R$
- $R + (S + T) \equiv (R + S) + T$
- $R . (S . T) \equiv (R . S) . T$
- $R . (S + T) \equiv (R . S) + (R . T)$
- $(S + T) . R \equiv (S . R) + (T . R)$
- $R + R \equiv R$
- $\varnothing^* \equiv \varepsilon$
- $R . R^* \equiv R^* . R \equiv R+$
- $R . R^* + \varepsilon \equiv R^*$

- **Exemples**. : *Expression régulière*

1. Le langage de ***tous les mots*** sur un alphabet A, A = $\{a_1, a_2, \ldots, a_n\}$

2. Le langage de *tous les mots non vide* sur un alphabet A = $\{a_1, a_2, \ldots, a_n\}$

3. Le langage de tous les mots *commençant* par ***a*** et se *terminant* par ***b*** sur l'alphabet A = $\{a, b, c\}$

4. L'ensemble des entiers naturels codés sur l'alphabet A=$\{0, 1, 2, \ldots, 9\}$

- **Exemples**. : *Expression régulière*

1. Le langage de ***tous les mots*** sur un alphabet quelconque A = $\{a_1, a_2, \ldots, a_n\}$ est dénoté par l'expression régulière : $\boldsymbol{(a_1 | a_2 | \ldots | a_n)^*}$. On écrit plus simplement **A***

1. Le langage de *tous les mots non vide* sur un alphabet quelconque A = $\{a_1, a_2, \ldots, a_n\}$ est dénoté par l'expression régulière $(a_1 | a_2 | \ldots | a_n)(a_1 | a_2 | \ldots | a_n)^*$. On écrit plus simplement **A.A*** ou $\mathbf{A}^+$

1. Le langage de tous les mots *commençant* par ***a*** et se *terminant* par ***b*** sur l'alphabet A = $\{a, b, c\}$ est dénoté par l'expression régulière $\boldsymbol{a(a|b|c)^*b}$

1. L'ensemble des entiers naturels codés sur l'alphabet
   $A=\{0, 1, 2, \ldots, 9\}$ dénoté par l'expression régulière suivante
   $0|(1|2| \ldots |9)(0|1|2| \ldots |9)^*$

Théorème :

Un langage est régulier si et seulement s'il est dénoté par une expression régulière.

Exercices

Exercice 1:

Soit l'alphabet A={a, b}.

Pour chacun des langages suivants, donner des exemples de mots, et donnez une expression régulière qui accepte chacun des langages

1. Le langage des mots contenant au moins une fois la lettre a :
2. Le langage des mots contenant au plus une fois la lettre a :
3. Le langage des mots contenant un nombre pair de fois la lettre a :
4. Le langage des mots admettant aba pour sous-mot

5. le langage des mots de longueur 2
6. L = { w {a,b}*/w= $a^n b^m a$ ou w = ba^n ; n,m ≥ 1 }
7. L= {w∈{a,b}*, tel que w contient seulement 3b, le reste c'est des a's}
8. L = {w∈{a,b}*, tel que w contient un nombre de ***a*** divisible par 3}
9. L = {w∈{a,b}*, tel que w contient un nombre *impaire* de b}
10. L= {w∈A*, w contient 3b consécutifs}
11. le langage des mots formés de ***n*** fois la lettre ***a*** suivi de ***n*** fois la lettre ***b***
12. le langage des mots comportant autant de ***a*** que de ***b***.

Exercice 2

1. Donner tous les mots de tailles 0, 1, 2, 3, et 4 des langages réguliers suivants :

   $L= (a + ba)^{*}$; $M=a(aa + b(ab)^{*} a)^{*}a$

Solution

Mots de longueurs 0 : ε ;

Mots de longueurs 1 : a ;

Mots de longueurs 2 : aa, ba ;

Mots de longueurs 3 : aaa, aba, baa ;

Mots de longueurs 4 : aaaa, aaba, abaa, baba, baaa.

- Mots de longueurs 0 : aucun ; Mots de longueurs 1 : aucun ; Mots de longueurs 2 : aa ; Mots de longueurs 3 : aucun ; Mots de longueurs 4 : aaaa, abaa.