Théories des langages

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

1

Notions fondamentales en Théorie des Langages

2

Introduction

- La théorie des Langages est une branche fondamentale de l'Informatique théorique.
- la première question que vous allez poser :
  - **Pourquoi faut-t-il étudier la théorie des Langages ?**
  - **C'est quoi l'apport pour un informaticien ?**

→Pour répondre à vos questions nous commençons ce chapitre par exposer des applications de cette théorie.

3

Exemple 1 : Correction d'orthographe et de grammaire dans un document Word

- Comment fonctionne un logiciel de correction d'orthographe ?
- Comment fonctionne un logiciel de correction de grammaire ?
- Si vous êtes entrain d'écrire un texte sur word, comment un mot peut être souligné en rouge ou en vert ?
- Quand est ce qu'un mot n'est pas souligné ?

Exemple2:Compilation d'un code avec le langage C

- Que se passe-t-il lorsque vous cliquez sur compiler ?
- Comment un compilateur vous présente les erreurs ?
- Le compilateur accepte tous les mots ?

4

INTRODUCTION

- Je suppose que vous posez déjà la question :

Y a-t-il une solution algorithmique pour ce genre de problèmes ?

→ la Théorie des Langages propose des solutions algorithmiques pour résoudre ces questions :

- **Y a-t-il une solution algorithmique permettant de vérifier : si un mot appartient à un Langage ?**

  → Pour répondre à cette question nous pouvons faire recours aux **automates**.

- **Y a-t-il une solution algorithmique permettant de vérifier : si une phrase respecte les règles de grammaire ?**

  →Pour répondre à cette question nous pouvons faire une étude des règles de **grammaire.**

LES LANGAGES

Définitions, propriétés et notations

Plusieurs notions sont fondamentales pour pouvoir présenter la théorie des langages.

- **Définition 1** (**Alphabet**) : Un alphabet, noté A, est un ensemble fini non vide de symboles (lettres, de chiffres et/ou de symboles graphiques)
  - ***Exemples d'alphabets :***

$$\begin{aligned} \mathcal{A}_1 &= \{ \bullet, \star, \diamond \} \\ \mathcal{A}_2 &= \{ \text{a, b, c, \ldots, z} \} \\ \mathcal{A}_3 &= \{ \text{if, then, else, id, nb}, =, + \} \end{aligned}$$

- **Définition 2** (**Mot**) : Un **mot** (ou bien une chaîne), défini sur un alphabet A, est une suite finie de symboles juxtaposés de A.
  - ***Exemples de mots :***
    - *sur l'alphabet* $\mathcal{A}_1$, *le mot* $\bullet\bullet\star$
    - *sur l'alphabet* $\mathcal{A}_2$, *le mot* if
    - *sur l'alphabet* $\mathcal{A}_3$, *le mot* if id $=$ nb
- **Définition 3** (**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***.
  - La cardinalité d'un mot ***u*** par rapport à un symbole $a \in$ A (notée par $|u|_a$) est le ***nombre d'occurrence*** de a dans ***u***.
  - $|abc| = 3$, $|1010111| = 7$.
  - *Le mot dont la longueur est nulle est noté par* ϵ : $|\epsilon| = 0$.
  - $|abc|_c = 1$, $|1010111|_1 = 5$, $|1010111|_0 = 2$.

7

- **Définition 4 (Mot vide)** : le mot vide, noté **ε**, est défini sur tous les alphabets et est le mot de longueur 0 (autrement dit, $|\varepsilon| = 0$).
- **Définition 5 (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.
- **Définition 6 (A^*)** : on note A^* l'ensemble des mots que l'on peut construire à partir de A, y compris le mot vide : $A^* = \{\varepsilon\} \cup A^+$

8

Concaténation des mots

Soient $\boldsymbol{u}$ et $\boldsymbol{v}$ deux mots définis sur l'alphabet $\mathcal{A}$. La concaténation de $\boldsymbol{u}$ avec $\boldsymbol{v}$ est un mot $\boldsymbol{w}$ défini sur le même alphabet $\mathcal{A}$. $\boldsymbol{w}$ est obtenu en écrivant $\boldsymbol{u}$ suivi de $\boldsymbol{v}$, en d'autres termes, on colle le mot $\boldsymbol{v}$ à la fin du mot $\boldsymbol{u}$:

- $\boldsymbol{u} = \boldsymbol{a}_1 \ldots \boldsymbol{a}_n$, $\boldsymbol{v} = \boldsymbol{b}_1\boldsymbol{b}_2 \ldots \boldsymbol{b}_m$
- $\boldsymbol{w} = \boldsymbol{uv} = \boldsymbol{a}_1 \ldots \boldsymbol{a}_n\boldsymbol{b}_1\boldsymbol{b}_2 \ldots \boldsymbol{b}_m$

- La concaténation est notée par le point, mais il peut être omis s'il n'y a pas de d'ambiguïté
- On écrira alors : $\boldsymbol{w} = \boldsymbol{u}.\boldsymbol{v} = \boldsymbol{uv}$
  - ***Exemples :***

9

Propriété de la concaténation

Soient $\boldsymbol{w}$, $\boldsymbol{u}$ et $\boldsymbol{v}$ trois mots définis sur l'alphabet A:

- $|\boldsymbol{u}.\boldsymbol{v}| = |\boldsymbol{u}| + |\boldsymbol{v}|$;
- $\forall \boldsymbol{a} \in A : |\boldsymbol{u}.\boldsymbol{v}|_{\boldsymbol{a}} = |\boldsymbol{u}|_{\boldsymbol{a}} + |\boldsymbol{v}|_{\boldsymbol{a}}$;
- $(\boldsymbol{u}.\boldsymbol{v}).\boldsymbol{w} = \boldsymbol{u}.(\boldsymbol{v}.\boldsymbol{w})$ (la concaténation est associative) ;
- $\boldsymbol{w}.\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}.\boldsymbol{w} = \boldsymbol{w}$ ($\boldsymbol{\varepsilon}$ est un élément neutre pour la concaténation) ;

10

L'exposant

L'opération $w.w$ est notée par w^2. En généralisant, on note $w^n = w...w$ (n fois)

En particulier, l'exposant **0** fait tomber sur **ε** : $w^0 = \varepsilon$ (le mot w est répété **0** fois).

Le mot miroir

Soit $w = a_1 a_2 ... a_n$ un mot sur A. On appelle mot miroir de w et on le note par w^R le mot obtenu en écrivant w l'envers, c'est-à-dire que $w^R = a_n ... a_2 a_1$. Il est donc facile de voir que $(w^R)^R = w$.

- Certains mots, appelés ***palindromes***, sont égaux à leur miroir ($w^R = w$).En d'autres termes, on lit la même chose dans les deux directions. Par ailleurs, on peut facilement vérifier que : $(u.v)^R = v^R.u^R$.

Définition (Préfixe, suffixe et facteur) :

Soient deux mots u et v définis sur un alphabet A.

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

Mots conjugués

Deux mots x et y sont dits conjugués s'il existe deux mots u et v tels que : $x = uv$ et $y = vu$.

Notions sur les langages

NOTIONS SUR LES LANGAGES

Un **langage** sur un alphabet A est un ***ensemble*** de mots construits sur A. Tout langage défini sur A est donc une partie de A*. On distingue en particulier le langage vide ϕ

Exemple :

- Langage des nombre binaires définies sur l'alphabet {0, 1} (infini) ;
- Langage des mots de longueur 2 défini sur l'alphabet {a, b}={aa, ab, ba, bb} ;
- Langage Pascal (quel est son alphabet ?) ;
- Langue française (quel est son alphabet ?).

Description d'un langage :

- Un langage ***fini*** peut être décrit par *l'énumération* des mots qui le composent.
- Certains langages ***infinis*** peuvent être décrits par *l'application d'opérations* à des langages plus simples.
- Certains langages ***infinis*** peuvent être décrits **par un ensemble de règles appelé grammaire**
- Enfin, certains langages infinis ne peuvent pas être décrits, ni par l'application d'opérations, ni par un ensemble de règles. On parle alors de langage indécidable.

Exemples .

1. Le langage des nombres binaires définies sur l'alphabet $\{0, 1\}$ (infini) ;
2. Le langage des mots de longueur 2 définis sur l'alphabet $\{0, 1\} = \{00, 01, 10, 11\}$ (fini)
3. La langue française définie sur l'alphabet $\{a, . . . , z\}$

« En tant qu'ensemble, les opérations ensemblistes restent applicables sur les langages. En plus, en profitant de la particularité des éléments constituant les langages : les mots, d'autres opérations non ensemblistes sont appliquées sur les langages. »

Opérations sur les langages:

Soient L, L_1 et L_2 trois langages définis sur l'alphabet A, nous définissons les opérations suivantes :

- **L'union** : notée par + ou | plutôt que $\cup$.

  $L_1 \mid L_2 = L_1 + L_2 = \{m \text{ tel que } m \in L_1 \vee m \in L_2\}$;
- **L'intersection** :

  $L_1 \cap L_2 = \{m \text{ tel que } m \in L_1 \wedge m \in L_2\}$;
- **Le complément** :

  $L = \{\text{tous les mots } m \text{ sur A } \textit{tel que } m \notin L\}$;
- La **concaténation** (opération non ensembliste) :

  $L1.L2 = \{m \text{ tel que } \exists u \in L1, \exists v \in L2 : m = uv\}$;
- **Exposant** (opération non ensembliste) :

  $L^n = L.L \ldots L$ (n fois)

  $= \{m \text{ tel que } \exists u_1, u_2, \ldots u_n \in L : m = u_1 u_2 \ldots u_n\}$

17

- **Fermeture transitive de Kleene** (opération non ensembliste) : notée $L^* = \cup_{i \geq 0} L^i$.

  En particulier, si L = A on obtient A*:

l'ensemble de tous les mots possibles sur l'alphabet A.

On peut ainsi définir un langage comme étant un sous-ensemble quelconque de A;

- **Fermeture non transitive** (opération non ensembliste) :

  $L^+ = \cup_{i>0} L^i$;
- **Le langage miroir** (opération non ensembliste) :

  $L^R = \{m \text{ tel que } \exists u \in L : m = u^R\}$

18

Propriétés des langages

Soit L, L_1, L_2, L_3 quatre langages définis sur l'alphabet A :

- $L^0 = \{\epsilon\}$
- $L^1 = L$
- $L^2 = L.L$
- $L^{n+1} = L.L^n = L^n.L$
- $L^* = L^+|\{\epsilon\}$;
- $L_1.(L_2.L_3) = (L_1.L_2).L_3$;
- $L_1.(L_2|L_3) = (L_1.L_2)|(L_1.L_3)$;
- $L.L \neq L$;
- $L_1.(L_2 \cap L_3) \neq (L_1 \cap L_2).(L_1 \cap L_3)$;
- $L_1.L_2 \neq L_2.L_1$;
- $(L^*)^* = L^*$;
- $L^*.L^* = L^*$;
- $L_1.(L_2.L_1)^* = (L_1.L_2)^*.L1$;
- $(L_1|L_2)^* = (L_1^*L_2^*)^*$;
- $L_1^*|L_2^* \neq (L_1|L_2)^*$;

- L'application des divers opérateurs doit respecter des priorités bien définies comme suit : | , . , *