Theorie des Langages

Matthew Coyle

January 10, 2018

1 Introduction

Les langages formels on ete etudies par:
- les informaticiens: langages de programmation
- les linguistes: langues naturelles

Langages formels: Points communs:
Chaque langages est un ensemble d'elements appeles mots ou "chaines".
Chaque chaine est une suite de symboles pris parmi un ensemble fini de symboles.
Chaque chaine est de longeur finie.

On etudie des modeles pour representer de maniere finie des langages:
automate finis
expressions refulieres
grammaire formelles

Application pratiques:
Recherche de "motifs" dans les fichiers
traitement de texte
modelisation de circuits
modelisation de machines a etats
Compliation de langages de programmation

un alphabet est un ensemble fini, non vide, de symoles. On le note generalement $\sum$.
Example d'alphabets:
$\sum_{entiers}$ = {0,1,2,3,4,5,6,7,8,9}
$\sum_{mots}$ = {a,b,c,...,z,',-}
$\sum_{ident}$ = {a,...,z,A,...,Z,0,...,9,_}
$\sum_{prog}$ = {int, float, bool, while, ¡, toto, a, ...}
U mot ou une chaine w forme(e) sur un alphabet est une suite finie $s_1 s_2 ... s_n$ de symboles de cet alphabet

La concatenation de deux chaines u et v, notee u.v ou uv est la chaine obtenue en ecrivant les symboles de u suivis de ceux de v.
si $u = a_1a_2...a_n$ et $v = b_1b_2b_p$
alors $uv = a_1a_2...a_nb_1b_2...b_p$

Un prefixe d'une chaine w est une suite de symboles debutant w.
Un suffixe de w est une suite de symboles terminant w.
$\forall x, y \mid w = x.y$, x est un prefixe de w, y un suffixe.

Une sous-chaine d'une chaine w est une suite de symboles apparaissant consecutivement dans w.

Un langage est un ensemble de chaine.
Example de langages:
{toto,titi,tata}
{1,11,101,1001}
$\{1^n \mid n \geq 0\} = \{e, 1, 11, 111, 1111, 11111, ...\}$
Nombres binaires impaires: {1,11,101,111,1001,1011,...}
Nombres binaires premiers: {1,10,11,101,111,1011,...}

le Langage vide, note ø, ne contient aucune chaine (ensemble vide).

Remaque: $\emptyset \neq \{\epsilon\}$
Le langage plein, note $\sum^*$

l'Union de deux langages A et B est le langage, note $A \cup B$, compose de toutes les chaines qui apparaissent dans l'un au moins de langages A ou B.

LA concatenation de deux langages A et B est le langage, note A.B ou AB, compose de toutes les chaines formees par une chaine de A concatenee a une chaine de B.
$A.B = \{u.v \mid u \in A$ ou $v \in B\}$

Proprietes:
Associativite: (A.B).C = A.(B.C)
$\{\epsilon\}$ est un element neutre: $A.\{\epsilon\} = \{\epsilon\}.A = A$
$\emptyset est element absorbant$
Distributivite de la concatenation sur l'union:

Puissance d'une langage A A^k est le langage forme par la concatenation de k occurrences de A.
$A^0 = \{\epsilon\}$
$A^1 = A$

Etoile de Kleene (fermeture ou cloture par *).

- la fermeture de Kleene d'un langage A est le langage, note A^*
$A^* = \bigcup|_{i=0}^{\infty} A^i$