THEORIE DES LANGAGES ET AUTOMATES

INTRODUCTION

- Les automates ont été définis dans les années 40-50.
- Ils sont utilisés dans les beaucoup de domaines comme:
  - Compilation des langages
  - Reconnaissance de texte
  - Conception de protocoles
  - Synthèse de programmes
  - Vérification de programmes
- Fin des années 50, Chomsky un linguiste a commencé l’étude des grammaires formelles.
- Les liens entre les grammaires et les automates ont été étudiés par la suite

- Alan Turing a étudié en 1930, avant que n’existe le premier ordinateur, une question fondamentale en Informatique:

Existe-t-il une limite à ce qu’on peut calculer?

Sa réponse:

oui, il y a des problèmes qu’on ne peut pas résoudre à l’aide d’un ordinateur même si on suppose une mémoire non bornée

- Bibliographie
  - P. Dehornoy, "Mathématiques d l’Informatique", Dunod, 2000.
  - J. Hopcroft et J. Ullmann, "Introduction to automata theory, Languages and Computations", Addison-Wesley, 1979.
- Webographie
  - http://njussien.e-constraints.net/lla/
  - http://brassens.upmf-grenoble.fr/~alecomte/

Plan

- Mots et langages
- Automates à états finis
- Langages réguliers
- Grammaires
- Langages hors contextes et automates à pile
- Machine de Turing

Mots et langages

Mots

- Un **alphabet** Σ est un ensemble fini non vide de lettres (ou symboles)

- Un **mot** est une suite finie de symboles de l'alphabet Σ

  Exemple: 001110 est un mot sur $\Sigma = \{0, 1\}$

- Un mot w constitué de m symboles est dit de longueur m. On note $|w| = m$.

  Exemple: $|001110|=6$

  $|\varepsilon|=0$

- On note Σ^* l'ensemble des mots sur Σ

- Σ^i est l'ensemble des mots de longueur i .
  On a : $\Sigma^* = \cup_n \Sigma^n$

- NB : On note Σ^+ l'ensemble $\Sigma^* \setminus \{\varepsilon\}$

Opérations

- **Concaténation**

  Soit x $\in\Sigma^*$ et y $\in\Sigma^*$ tels que $|x| = m$ et $|y| = n$.

  La concaténation de x et y notée xy est le mot de longueur m + n dont les m premiers symboles sont ceux de x et les n derniers ceux de y

- Propriétés de la concaténation
  - associativité
  - non commutativité
  - ε neutre pour la concaténation

- Portions de mots:

  Soient w, x, y, z quatre mots tels que : w = xyz

  - x est un **préfixe** de w
  - y est une **sous-chaîne** de w
  - z est un **suffixe** de w

- Exemple

  Si le mot considéré est w = 00110

  - préfixes : ε, 0, 00, 001, 0011 et w
  - suffixes : ε, 0, 10, 110, 0110 et w
  - sous-chaînes : ε, 0, 1, 00, 01, 10, 11, 001, 011, 110, 0011,0110 et w

- L'opération **miroir** est définie de la manière suivante :
  - si $|w| = 0$, alors $w = \varepsilon$ et $w^R = \varepsilon$
  - Sinon $|w| > 0$, $w = au$ ($a \in \Sigma$ et $u \in \Sigma^*$), $w^R = u^R a$.

Si w est tel que $w^R = w$ alors w est un **palindrome.**

Langages

- **Langage**

Un langage (formel) L sur un alphabet Σ est un sous-ensemble quelconque de Σ^* tel que $L \subseteq \Sigma^*$

- Exemple

Le langage des nombres est défini sur un alphabet $\Sigma = \{0, 1, . . . , 9\}$. 02, 00310, 3200 sont alors des mots sur Σ. On définira le langage des nombres comme les mots sur Σ qui ne commencent pas par 0. Ainsi, 1233 et 3200 seront des mots du langages mais pas 00310.

- **Concaténation de langages**

  Soient L1 et L2 deux langages, leur concaténation est le langage

  $L1L2 = \{xy | x \in L1, y \in L2\}$

- Propriétés de la concaténation
  - associative
  - non commutative
  - $\{\varepsilon\}$ neutre
  - $\varnothing$ absorbant

- **Clôture de Kleene**
  - $L^2 = LL$ la concaténation de L avec lui-même
  - $L^0 = \{\varepsilon\}$
  - clôture de Kleene

$$L^* = \cup_{i=0 \to \infty} L^i$$

- Autres opérations sur les langages:
  - $A + B$ (ou $A \cup B$) = $\{w \in X^* \mid w \in A$ ou $w \in B\}$
  - $A \cap B = \{w \in X^* \mid w \in A$ et $w \in B\}$
  - $A - B$ (ou $A \setminus B$) = $\{w \in X^* \mid w \in A$ et $w \notin B\}$
  - $L^R=\{u^R \mid u \in L\}$ (image miroir d'un langage)

- Propriétés:
  - $A \emptyset = \emptyset A = \emptyset$
  - $A \{\varepsilon\} = \{\varepsilon\} A = A$
  - $A \subseteq B \Rightarrow AC \subseteq BC$
  - $A^* = \cup_{i=0 \to \infty} A^i$
  - $A^+ = \cup_{i=1 \to \infty} A^i$

- Remarque:

  On n' a pas la distributivité de la concaténation par rapport à l'intersection

Théorème d'Arden

A, B langages sur Σ.

Si $\varepsilon \notin A$ alors l'équation $L=AL+B$ admet une unique solution $L= A^*B$