IN 406 – Théorie des Langages
Cours 1 : Mot, langage et automate

Franck Quessette – Franck.Quessette@uvsq.fr

Université de Versailles – Saint-Quentin

V3 2019–2020

Organisation de l'UE

- 12 Cours et 12 TD ;
- Franck Quessette pour les huit premiers cours ;
- Guillaume Scerri pour les 4 derniers ;
- 5 groupes de TD :
  - Gr 3 : lundi 13h40-15h10 Amphi I Xavier Badin de Montjoye ;
  - Gr 1, 2 et DL : mardi 09h40-12h50 Salles G002, RC22, D101 Loric Duhazé, Franck Quessette, Yann Strozecki ;
  - Gr 4 : mardi 13h50-15h10 Salle G210 Guillaume Scerri.
- 1 note de CC (1/3) et une note d'examen (2/3) ;
- CC : projet + contrôle.

Objectifs de l'UE

Définir et montrer l'utilisation d'outils mathématiques :

- alphabet, mot, langage ;
- expression régulière ;
- grammaire ;
- automate, automate à pile ;
- machine de Turing.

Introduction aux modèles de calculs :

- machine de Turing ;
- calculabilité.

Alphabet, mot

Définition

Un $\boxed{\textbf{alphabet}}$ est un ensemble fini de symboles, appelés lettres ou caractères. Cet ensemble est généralement noté Σ (sigma).

Exemples

- $\Sigma_1 = \{a, b\}$, un alphabet à deux lettres ;
- $\Sigma_2 = \{a_1, a_2, \ldots, a_n\}$, un alphabet à n lettres.

Définition

Un $\boxed{\textbf{mot}}$ sur un alphabet Σ est une concaténation finie de lettres de Σ.

Exemples

Soit $\Sigma = \{a, b, c\}$ un alphabet,

- $w_1 = abc$, est un mot de trois lettres. $w_2 = aabaa$, est un mot de cinq lettres, chaque lettre peut apparaître plusieurs fois ;
- $w_1 = abc$ et $w_3 = acb$ sont deux mots différents, la concaténation n'est pas commutative.
- La concaténation est parfois notée : $w = a \cdot b \cdot c$.
- $w_4 = c$ est un mot.

Attention a représente à la fois la lettre a et le mot d'une seule lettre a qui est parfois notée "a".

Concaténation

La concaténation peut être définie entre deux lettres, entre deux mots ou entre un mot et une lettre :
Soient $\Sigma = \{a, b, c\}$ un alphabet, et $w_1 = abc$ et $w_2 = aabaa$ des mots sur l'alphabet Σ alors,
$w_1 c = abcc$, $bw_1 = babc$ et $w_1 w_2 = abcaabaa$ sont des mots sur l'alphabet Σ.

Définition

La **taille** ou **longueur** d'un mot w est le nombre de lettres qui le composent, ce nombre est noté $|w|$. De plus, la notation $|w|_a$ est le nombre d'occurences de la lettre a dans le mot w.

Exemple

Si $w = aabaa$ est un mot sur l'alphabet $\Sigma = \{a, b, c\}$, alors
$|w| = 5$, $|w|_a = 4$, $|w|_b = 1$ et $|w|_c = 0$.

Préfixe, suffixe

Définition

Un mot w_1 est un $\boxed{\textbf{préfixe}}$ d'un mot w s'il existe un mot w_2 tel que $w = w_1 w_2$.
Un mot w_2 est un $\boxed{\textbf{suffixe}}$ d'un mot w s'il existe un mot w_1 tel que $w = w_1 w_2$.

Définition

Le caractère $\boxed{\varepsilon}$ (epsilon) est le caractère vide, il est tel que pour tout mot w, $w\varepsilon = \varepsilon w = w$.
ε représente également le mot vide : $|\varepsilon| = 0$.

Exemple

Si $w = aabc$, les préfixes de w sont : ε, a, aa, aab et $aabc$ et les suffixes de w sont : $aabc$, abc, bc, c et ε.

Notations

- Pour simplifier l'écriture, un mot composé de caractères identiques aaa peut être noté a^3.
- De même pour des sous-mots, $w = aabbaaa = a^2 b^2 a^3$.
- Généralisation aux mots, si $w = abc$, alors $w^2 = abcabc$, $w^1 = w$ et $w^0 = \varepsilon$.

Langage

Définition

Un $\boxed{\textbf{langage}}$ sur un alphabet Σ est un ensemble fini ou infini de mots.

Exemples

Soit l'alphabet $\Sigma = \{a, b, c\}$

- $L_0 = \emptyset$, ce langage ne contient aucun mot ;
- $L_1 = \{\varepsilon\}$, ce langage contient un mot qui est le mot vide, donc $L_1 \neq L_0$;
- $L_2 = \{abc, bca, bbb\}$, ensemble fini ;
- $L_3 = \{a^n, n \geq 0\}$, ensemble infini ;
- L_4 l'ensemble des mots qui contiennent un nombre pair de c.

Un langage peut-être défini en français, par une formule ou bien à l'aide d'un formalisme mathématique : expression régulière, grammaire, automate, ...

Différents types de problèmes

Problème de **calcul** :

Exemples

- multiplier deux nombres ;
- factoriser deux nombres.

Problème d' **optimisation** :

Exemples

- calculer le plus court chemin d'ici à là ;
- découper des pièces dans un morceau de tissu pour minimiser les chutes ;
- affecter les personnels navigants sur les avions en minimisant les coûts.

Problème de décision

Problème de **décision** : question avec une réponse OUI/NON.

Exemples

- ce graphe est-il planaire ?
- ce nombre est-il premier ?
- cette machine de Turing s'arrête-t-elle toujours ?
- ce mot appartient-il à ce langage ?

Tous les problèmes peuvent se ramener à des problèmes de décision ou au moins à une suite de problèmes de décisions.

QUESTION FONDAMENTALE :
Soit L un langage et w un mot, est-ce que $w \in L$?

Réponse automatique à la question fondamentale

QUESTION FONDAMENTALE :
Soit L un langage et w un mot, est-ce que $w \in L$?

Peut-on construire une machine universelle qui répond à la question fondamentale pour tout langage et tout mot ?

La réponse est NON,
il n'existe pas de machine universelle.

Ce que l'on va faire dans ce cours :

Classification des langages et des machines permettant de répondre à la question fondamentale.

Concaténation de langages

Définition

La **concaténation de deux langages** L_1 et L_2 est notée $L = L_1 L_2$ est définie par :

$$w \in L \iff \exists w_1 \in L_1, \exists w_2 \in L_2, \text{tels que } w = w_1 w_2$$

Exemples

- $L = \{ab, b\}$, $LL = \{abab, abb, bab, bb\}$
- $L_1 = \{ab, b\}$, $L_2 = \{bb, ba\}$, $L_1 L_2 = \{abbb, abba, bbb, bba\}$.
- $L_1 = \{ab, a\}$, $L_2 = \{ba, a\}$, $L_1 L_2 = \{abba, aba, aa\}$.

Puissance d'un langages

Définition

La **puissance** d'un langage L est définie par :

- $L^0 = \{\epsilon\}$;
- $L^n = LL^{n-1}$ pour $n > 0$.

Par induction, $L^1 = L$.

Exemples

- $L = \{ab, b\}$, $LL = L^2 = \{abab, abb, bab, bb\}$

Définition

L' $\boxed{\textbf{étoile de Kleene}}$ d'un langage L, notée L^* est définie par :

$$L^* = \bigcup_{n=0}^{+\infty} L^n$$

Exemples

Soit l'alphabet $\Sigma = \{a, b\}$

- $L = \{a\}$, $L^* = \{a^n, n \geq 0\}$;

Un alphabet Σ peut être vu comme un langage ne contenant que des mots d'une lettre et avec $\Sigma = \{a, b\}$:
$\Sigma^* = \{\varepsilon, a, b, aa, ab, ba, bb, \ldots\}$, l'ensemble de tous les mots possibles sur l'alphabet Σ.

Opérations sur les langages

Les langages étant des ensembles, les opérations ensemblistes (union, intersection, complémentaire) sont naturellement définies.

Exemples

Soit l'alphabet $\Sigma = \{a, b, c\}$, L_1 l'ensemble des mots qui commencent par a ou b et L_2 l'ensemble des mots qui commencent par b ou c.

- **Intersection** : $L_1 \cap L_2$ est l'ensemble des mots qui commencent par b.
- **Union** :$L_1 \cup L_2$ est l'ensemble de **tous** les mots sauf ε.
- **Complémentaire** : $\overline{L_1} = \Sigma^* \backslash L_1$ est l'ensemble des mots qui commencent $c \cup \{\varepsilon\}$.

Définition

Un $\boxed{\textbf{langage rationnel}}$ sur un alphabet Σ est défini par :

- $L = \emptyset$ est rationnel ;
- $L = \{\varepsilon\}$ est rationnel ;
- pour tout $a \in \Sigma$, $L = \{a\}$ est rationnel ;
- si L_1 et L_2 sont rationnels alors $L_1 L_2$ et $L_1 \cup L_2$ sont rationnels ;
- si L est rationnel alors L^* est rationnel.

Lemme

Tout langage fini est rationnel.

Première machine : automate fini

Définition

Un **automate fini** $\mathcal{A}$ est un quintuplet $\mathcal{A} = (\Sigma, Q, q_0, F, T)$ avec :

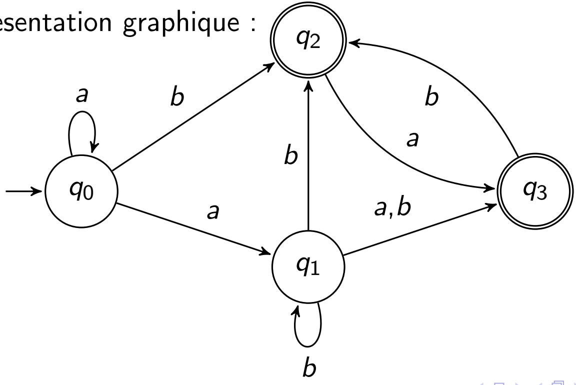
- Σ un alphabet fini ;
- Q un ensemble fini d'états ;
- $q_0 \in Q$ l'état initial ;
- $F \subseteq Q$ l'ensemble des états finaux ;
- $T \subseteq Q \times \Sigma \times Q$ l'ensemble des transitions.

Exemple

$\mathcal{A} = (\Sigma, Q, q_0, F, T)$ avec $\Sigma = \{a, b\}$, $Q = \{q_0, q_1, q_2, q_3\}$,
$F = \{q_2, q_3\}$,
$T = \{(q_0, a, q_0), (q_0, a, q_1), (q_0, b, q_2), (q_1, b, q_1), (q_1, b, q_2), (q_1, a, q_3), (q_1, b, q_3), (q_2, a, q_3), (q_3, b, q_2)\}$.

Représentation graphique :

$\mathcal{A}$:

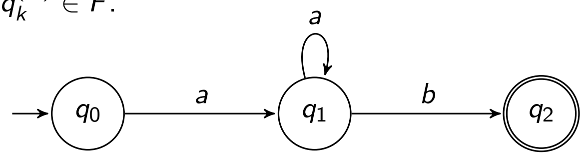


Définition

Un automate fini $\mathcal{A} = (\Sigma, Q, q_0, F, T)$ **reconnait** le mot $w = a_1 a_2 \dots a_k$, s'**il existe** une séquence d'états $q_0^{(w)}, q_1^{(w)}, \dots, q_k^{(w)}$ telle que :

- $q_0^{(w)} = q_0$;
- $\forall i \in 0..k-1,\ (q_i^{(w)}, a_{i+1}, q_{i+1}^{(w)}) \in T$;
- $q_k^{(w)} \in F$.

$\mathcal{A}$:



- *aaaab* reconnu avec la séquence d'états $q_0, q_1, q_1, q_1, q_1, q_2$.
- *aba* non reconnu, pas de transition pour le deuxième a.
- *aa* non reconnu, arrêt dans un état non final.

Reconnaissance d'un langage

Définition

Le $\boxed{\textbf{langage reconnaissable}}$ par un automate fini $\mathcal{A}$ est noté $L(\mathcal{A})$ et est défini par :

$$w \in L(\mathcal{A}) \Longleftrightarrow w \text{ est reconnu par } \mathcal{A}$$

Attention à l'équivalence dans la définition.

Théorème de Kleene

Théorème de Kleene

Pour tout langage L,

$$L \text{ est } \boxed{\textbf{reconnaissable}} \iff L \text{ est } \boxed{\textbf{rationnel}}.$$

La preuve est faite en TD.

Voir également : fr.wikipedia.org/wiki/Théorème_de_Kleene.