

Langages, Compilation, Automates.

Partie 1: alphabets, mots, langages, automates finis deterministes.

Florian Bridoux

Polytech Nice Sophia

2022-2023

Table des matières

- 1 Organisation du cours
- 2 Introduction:
- 3 Alphabets, mots et langages
- 4 Automates finis déterministes
- 5 Compléter un automate
- 6 Le complémentaire d'un langage rationnel

- 1 Organisation du cours
- 2 Introduction:
- 3 Alphabets, mots et langages
- 4 Automates finis déterministes
- 5 Compléter un automate
- 6 Le complémentaire d'un langage rationnel

- Équipe enseignante:
 - Florian Bridoux (CM et TD),
 - Igor Litovsky (TD)
 - Hélène Renard Ferry (TD)
- Séances:
 - 13 CM de 1h.
 - 13 TD de 2h30
- Évaluation (à confirmer):
 - Projet en binôme: réaliser un compilateur
 - Tests Moodle notés.
 - Examen terminal de 2h sur papier (documents autorisés).

Table des matières

- 1 Organisation du cours
- 2 Introduction:
- 3 Alphabets, mots et langages
- 4 Automates finis déterministes
- 5 Compléter un automate
- 6 Le complémentaire d'un langage rationnel

Les langages permettent aux Hommes de donner des instructions et d'échanger des informations entre eux avec leurs machines.

- Les langages utilisés dans la vie de tous les jours entre êtres humains (le français, anglais, japonais, etc.) sont dits naturels. Ils sont généralement informels et ambigus et demandent toute la subtilité d'un cerveau humain pour être interprétés correctement.
- Les langages créés par l'homme pour communiquer avec les ordinateurs (python, C++, Java, langage assembleur, ...) sont des langages artificiels. Ils doivent être formalisés et non ambigus pour pouvoir être interprétés par une machine.

(Concernant les langages naturels, petite parenthèse sur les correcteurs d'orthographe et de grammaire comme **Grammalecte**)

Objectif du cours:

- Savoir formaliser des langages formels. D'abord des langages "simples" (dit **régulier**), puis des langages plus complexes.
- Au départ, un ordinateur ne comprend pas le python, le C++, etc. Pour communiquer avec l'ordinateur, il est nécessaire de transformer le code dit haut niveau en un langage machine. Cette tâche est réalisée par un programme appelé compilateur (ou un interpréteur). À la fin du cours, vous devriez savoir créer un compilateur *basique* de A à Z.

Table des matières

- 1 Organisation du cours
- 2 Introduction:
- 3 Alphabets, mots et langages**
- 4 Automates finis déterministes
- 5 Compléter un automate
- 6 Le complémentaire d'un langage rationnel

Definition (Alphabet)

Un **alphabet**, souvent noté Σ , est un ensemble fini non vide d'éléments appelés symboles (ou lettres).

Quelques exemples:

- $A = \{0, 1\}$: l'alphabet binaire.
- $B = \{a, b, c\}$: un alphabet composé des lettres a et b et c .

Definition (Mot)

Un **mot** est une suite **finie** de symbole sur un alphabet donné.

Sur l'alphabet A , 0011, 10001 et 0000 sont des mots. Sur l'alphabet B , abc , $cccc$ et $abba$ sont des mots.

Definition (Longueur d'un mot)

La **longueur** d'un mot u notée $|u|$ est son nombre de symboles.

Par exemple, sur l'alphabet $\{a, b\}$, pour les mots $u = a$, $v = abba$ et $w = bbbb$, on a $|u| = 1$, $|v| = 4$ et $|w| = 4$.

Definition (Le mot vide ϵ)

Le **mot vide**, noté ϵ est le seul mot de longueur nulle ($|\epsilon| = 0$).

Alphabets, mots et langages

Pour tout alphabet Σ et pour toute longueur $\ell \in \mathbb{N}$, on note Σ^ℓ l'ensemble des mots de longueur ℓ sur l'alphabet Σ .

Exercice

Pour $\Sigma = \{a, b\}$, que vaut Σ^3 ?

Exercice

Soit Σ un alphabet et ℓ un entier. Exprimez $|\Sigma^\ell|$ en fonction de $|\Sigma|$ et ℓ .

On note Σ^* l'ensemble des mots sur l'alphabet Σ . On a donc

$$\Sigma^* = \bigcup_{\ell \in \mathbb{N}} \Sigma^\ell.$$

On note également $\Sigma^+ = \Sigma^* \setminus \{\epsilon\}$.

Definition (Langage)

Un **langage** L sur l'alphabet Σ est un ensemble de mots sur cet alphabet: $L \subseteq \Sigma^*$.

Quelques exemples de langages:

- le langage vide $|L| = 0$, le langage composé du mot vide $L = \{\epsilon\}$, le langage composé de tous les mots sur un alphabet donné $L = \Sigma^*$,
- sur l'alphabet de la langue française, l'ensemble des mots défini dans la dernière édition du Larousse.
- sur l'alphabet binaire, l'ensemble des nombres binaires correctement formés: $\{0, 1, 10, 11, \dots\}$.
- sur l'alphabet décimal $\Sigma = \{0, 1, 2, \dots, 9\}$:
 - l'ensemble des nombres multiples de 3: $L = \{0, 3, 6, 9, \dots\}$,
 - l'ensemble des nombres premiers: $\{2, 3, 5, 7, 11, \dots\}$.

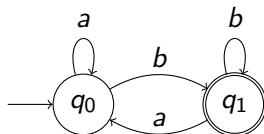
Table des matières

- 1 Organisation du cours
- 2 Introduction:
- 3 Alphabets, mots et langages
- 4 Automates finis déterministes**
- 5 Compléter un automate
- 6 Le complémentaire d'un langage rationnel

Definition (Automate fini déterministe (AFD))

Un **automate fini déterministe (AFD)** est un quintuplet $(\Sigma, Q, \delta, q_0, F)$ où:

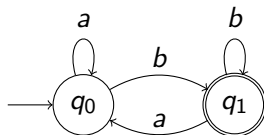
- Σ est un alphabet,
- Q est un ensemble **fini** d'états,
- δ est la fonction de transition: $Q \times \Sigma \rightarrow Q$,
- $q_0 \in Q$ est l'état initial,
- $F \subseteq Q$ est l'ensemble des états d'acceptation.



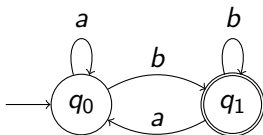
Automates finis déterministe

Un automate fini déterministe A est une machine qui calcule si un mot w appartient au langage $L(A)$ défini par cet automate.

- L'automate commence dans l'état q_0 .
- L'automate lit les lettres du mot de gauche à droite.
- Si l'automate est dans l'état q_i , qu'il lit une lettre ℓ , et que $\delta(q_i, \ell) = q_j$ alors l'automate passe dans l'état q_j (on écrit alors $q_i \xrightarrow{\ell} q_j$).
- Si $\delta(q_i, \ell)$ n'est pas défini alors l'automate rejette le mot (on écrit alors $q_i \xrightarrow{\ell} \perp$).
- Après avoir lu tout le mot, l'automate accepte le mot si et seulement s'il est dans un état acceptant.



Automates finis déterministe



Sur le mot *abba*, l'automate va effectuer les transitions suivantes:

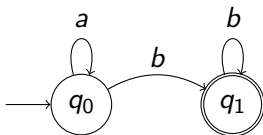
$$q_0 \xrightarrow{a} q_0 \xrightarrow{b} q_1 \xrightarrow{b} q_1 \xrightarrow{a} q_0.$$

L'automate va s'arrêter dans un état $q_0 \notin F$ et va donc rejeter le mot. Donc $L(A)$ ne contient pas le mot *abba*.

Exercice

- Écrire la liste des transitions prises par l'automate pour les mots ϵ , *bab* et *baba*.
- Lesquels de ces mots appartiennent à $L(A)$?
- Expliciter le langage $L(A)$.

Automates finis déterministe



Sur le mot *abba*, l'automate va effectuer les transitions suivantes:

$$q_0 \xrightarrow{a} q_0 \xrightarrow{b} q_1 \xrightarrow{b} q_1 \xrightarrow{a} \perp.$$

L'automate va donc rejeter le mot. Donc $L(A)$ ne contient pas le mot *abba*.

Exercice

- Écrire la liste des transitions prises par l'automate pour les mots *babbb*, *abab* et *aabbb*.
- Lesquels de ces mots appartiennent à $L(A)$?
- Expliciter le langage $L(A)$.

Exercice

Écrire un automate qui sur l'alphabet $\{a, b, c\}$ reconnait les mots contenant le facteur cba .

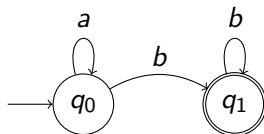
Table des matières

- 1 Organisation du cours
- 2 Introduction:
- 3 Alphabets, mots et langages
- 4 Automates finis déterministes
- 5 Compléter un automate**
- 6 Le complémentaire d'un langage rationnel

Compléter un automate

Definition (Automate complet)

Un automate $A = (\Sigma, Q, \delta, q_0, F)$ est **complet** si pour tout $q \in Q$ et tout $\ell \in \Sigma$, $\delta(q, \sigma)$ est bien défini.



Exercice

Sur l'alphabet $\{a, b\}$, trouver un automate **complet** qui reconnaît le même langage que l'automate ci-dessus.

Exercice

D'une manière générale, décrire la procédure pour transformer un automate non-complet en un automate complet sans changer le langage reconnu.

Table des matières

- 1 Organisation du cours
- 2 Introduction:
- 3 Alphabets, mots et langages
- 4 Automates finis déterministes
- 5 Compléter un automate
- 6 Le complémentaire d'un langage rationnel**

Compléter un automate

Definition (Langage rationnel)

Un **langage rationnel** est un langage qui est reconnu par un AFD.

Autrement dit, L est un langage rationnel si et seulement s'il existe un AFD A tel que $L = L(A)$.

Definition (Le complémentaire)

Le complémentaire \bar{L} d'un langage L sur un alphabet Σ est $\Sigma^* \setminus L$.

Par exemple, le complémentaire sur l'alphabet $\Sigma = \{a\}$ de $L = \{a^n \mid n \text{ est impair} \}$, $\bar{L} = \{a^n \mid n \text{ est pair} \}$.

Théorème

Le complémentaire d'un langage rationnel est un langage rationnel.

Exercice

Prouver ce théorème.