

IFT-2002

Informatique Théorique

H14 - cours 4

Julien Marcil - julien.marcil@ift.ulaval.ca

Langage
non régulier

Observation

Soit le langage $L = \{w \mid w \text{ contient le même nombre de } 0 \text{ et de } 1\}$. Est-ce que L est régulier?

Lemme de pompage

Si L est un langage régulier, alors il existe un entier $p \geq 1$ (appelé longueur de pompage) tel que pour tout mot $w \in L$ avec $|w| \geq p$, il existe des mots x, y, z tels que $w = xyz$ et

1. $|xy| \leq p$
2. $|y| > 0$
3. pour tout entier $i \geq 0$ on a $xy^iz \in L$

Preuve qu'un langage L est non régulier

Pour prouver qu'un langage est non régulier, on fait une preuve par contradiction.

- On suppose que L est régulier.
- Donc il existe p la longueur de pompage de L .
- On *choisi* un mot $w \in L$, avec $|w| \geq p$ (qu'il ne sera pas possible de pomper).
- On montre qu'en pompant w , on génère des mots qui ne sont pas dans L .

Exemple

Prouver que le langage $L = \{0^n 1^n \mid n \in \mathbb{N}\}$ n'est pas régulier.

Exemple

Prouver que le langage

$L = \{w \mid w \text{ contient le même nombre de 0 et de 1}\}$ n'est pas régulier.

Exemple

Prouver que le langage $L = \{0^i 1^j \mid i > j\}$ n'est pas régulier.

Exemple

Prouver que le langage $L = \{0^i 1^j \mid i \neq j\}$ n'est pas régulier.

Exemple

Prouver que le langage $L = \{1^{n^2} \mid n \in \mathbb{N}\}$ n'est pas régulier.

$$L = \{1, 1111, 111111111111, \dots\}$$

Exemple

Est-ce que le langage $L = \{w \mid w \text{ contient le même nombre de sous-chaînes } 01 \text{ et } 10\}$ est régulier?

Exemple

Soit $\Sigma = \{0, 1\}$ et

$$A = \{0^k u 0^k \mid k \geq 1, u \in \Sigma^*\}$$

$$B = \{0^k 1 u 0^k \mid k \geq 1, u \in \Sigma^*\}$$

- Montrer que A est régulier
- Montrer que B est non régulier

Grammaire

Introduction

Dans cette section, nous verrons un nouvel outil servant à la *génération* de langages: **les grammaires**.

Changement de point de vue: au lieu de considérer des machines qui *reconnaissent* un langage donné, on construit un ensemble de règles qui servent à *construire* les mots d'un langage.

Grammaire

Une grammaire consiste en un ensemble fini de **règles de réécriture** (ou *règles de productions* ou *règles de substitution*).

$$S \rightarrow 0S1$$

$$S \rightarrow B$$

$$B \rightarrow \lambda$$

Dérivation

Une grammaire *génère* une séquence par **dérivation**.

$$\begin{aligned} S &\Rightarrow 0S1 \\ &\Rightarrow 00S11 \\ &\Rightarrow 000S111 \\ &\Rightarrow 000B111 \\ &\Rightarrow 000\lambda 111 \\ &\Rightarrow 000111 \end{aligned}$$

Grammaire

Définition: Une **grammaire** consiste en un quadruplet de la forme (V, Σ, S, R) où

- V est ensemble fini de *variables* (symboles non terminaux).
- Σ est l'*alphabet* (symboles terminaux).
- $S \in V$ est le *symbole de départ*.
- R est un ensemble fini de *règles de réécriture*.

Règles de réécriture

Ces règles sont formées d'un terme de gauche, d'une flèche (\rightarrow) et d'un terme de droite.

Les termes gauche et droit peuvent être n'importe quelle combinaison de symboles de V ou de Σ , pourvu qu'il y ait au moins un symbole de V à gauche. Le côté droit peut être vide, ce qui est indiqué par un λ .

Exemple

Voici une grammaire (V, Σ, S, R) où

$$V = \{S, A, B, C\}$$

$$\Sigma = \{a, b, c\}$$

S est le symbole de départ

$$R = \{ S \rightarrow ASC$$

$$S \rightarrow B$$

$$B \rightarrow bB$$

$$B \rightarrow \lambda$$

$$A \rightarrow a$$

$$C \rightarrow c \}$$

Conventions

- Les majuscules sont des symboles non terminaux.
- Les minuscules sont des symboles terminaux.
- S est le symbole initial.

Cette convention permet de simplifier la description d'une grammaire en donnant seulement la description de R .

Notation

$$\begin{array}{ll} S \rightarrow ASC & S \rightarrow B \\ B \rightarrow bB & B \rightarrow \lambda \\ A \rightarrow a & C \rightarrow c \end{array}$$

Il est possible de noter les règles de réécriture plus simplement

$$\begin{array}{l} S \rightarrow ASC \mid B \\ B \rightarrow bB \mid \lambda \\ A \rightarrow a \\ C \rightarrow c \end{array}$$

Production

On nomme **production** l'application d'une règle de réécriture à une chaîne de symboles terminaux et non terminaux.

Soit $u, v, w \in (V \cup \Sigma)^*$, et la règle de réécriture $A \rightarrow w$.
On dit que uAv produit uwv , noté

$$uAv \Rightarrow uwv$$

Dérivation

On nomme **dérivation** l'application d'une ou plusieurs règles de réécriture à une chaîne de symboles terminaux et non terminaux.

Soit $u, v \in (V \cup \Sigma)^*$ on dit u dérive v , noté

$$u \xRightarrow{*} v$$

si $u = v$ ou s'il existe u_1, u_2, \dots, u_k pour $k \geq 0$ et

$$u \Rightarrow u_1 \Rightarrow u_2 \Rightarrow \dots \Rightarrow u_k \Rightarrow v$$

