Langages et Automates
Automates et lemme de l'étoile

Engel Lefaucheux

Prépas des INP

Objectif du cours

- Mieux comprendre le lemme de l'étoile
- Minimiser un automate

Lemme de l'étoile

Theorem (Théorème de Kleene)

Les langages reconaissables sont exactement les langages réguliers.

Theorem

Soit L un langage régulier. Il existe un entier N tel que tout mot w de L de longueur $|w| \geq N$ possède une factorisation $w = xyz$ avec $0 < |y|$ telle que

1. $0 < |xy| \leq N$ *et*
2. $xy^n z \in L$ *pour tout entier* $n \geq 0$.

Lemme de l'étoile

Theorem (Théorème de Kleene)

Les langages reconaissables sont exactement les langages réguliers.

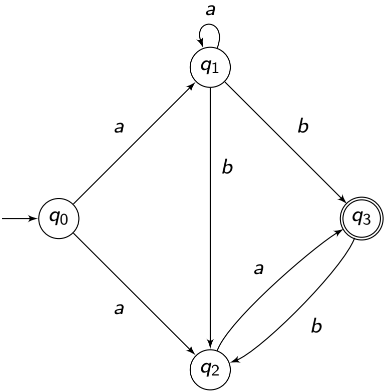
Theorem

Soit L un langage régulier. Il existe un entier N tel que tout mot w de L de longueur $|w| \geq N$ possède une factorisation $w = xyz$ avec $0 < |y|$ telle que

1. $0 < |xy| \leq N$ *et*
2. $xy^n z \in L$ *pour tout entier* $n \geq 0$.

Que signifie ce lemme, du point de vue de l'automate ?

Exemple



Theorem

Soit L un langage régulier. Il existe un entier N tel que tout mot w de L de longueur $|w| \geq N$ possède une factorisation $w = xyz$ avec $0 < |y|$ telle que

1. $0 < |xy| \leq N$ *et*
2. $xy^nz \in L$ *pour tout entier* $n \geq 0$.

Soit $N = 5$

Donner un mot de ce langage de longueur supérieur ou égal à 5

Preuve formelle

Soit $\mathcal{L}$ un langage régulier et $\mathcal{A} = (Q, \Sigma, T, I, F)$ l'automate reconnaissant ce langage. Soit N le nombre d'états de $\mathcal{A}$.

Soit $w = w_1 \ldots w_m \in \mathcal{L}$ avec $m \geq N + 1$. Comme w est accepté par $\mathcal{A}$, il existe un chemin

$$q_0 \xrightarrow{w_1} q_1 \xrightarrow{w_1} q_2 \xrightarrow{w_2} \ldots \xrightarrow{w_m} q_m$$

où $q_0 \in I$ et $q_m \in F$.
Comme $m \geq N + 1$, par le principe des tiroirs, il existe $1 \leq i < j \leq N + 1$ tel que $q_i = q_j$. On fixe $x = w_1 \ldots w_i$, $y = w_{i+1} \ldots w_j$ et $z = w^{j+1} \ldots w_m$. Au mot $xy^k z$ correspond le chemin acceptant

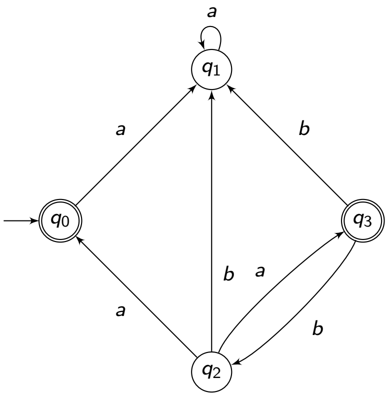
$$q_0 \xrightarrow{w_1} q_1 \xrightarrow{w_1} \ldots \xrightarrow{w_i} q_i (\xrightarrow{w_{i+1}} \ldots \xrightarrow{w_j} q_j)^k \xrightarrow{w_{j+1}} \ldots \xrightarrow{w_m} q_m$$

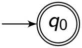
Exercice

Donner la constante N du lemme de l'étoile associée aux langages suivants

- $a^* b^*$
- $a + bb(a + b)^*$
- $a^4 b^3$
- $b^* aa(b + a)$

Minimiser un automate





Ces deux automates ont le même langage : $\{\varepsilon\}$

Quel algorithme pour minimiser un automate ?

Langage résiduel

Soit L un langage, $w \in \Sigma^*$.
Le résiduel de L par w est

$$w^{-1}L = \{v \in \Sigma^* \mid wv \in L\}$$

Exemple : $(aa)^{-1}((a+b)^2 b^* = b^*$

Exercice

Calculer

- $(aa)^-1(b^*ab^*ab^*)$
- $(\varepsilon)^-1(\varepsilon)$
- $(aa)^-1(b(a+b)^*)$

Calculer l'ensemble des résiduels possibles des langages

- a^*b^*
- $aa(ba)^*$

Automate des résiduels

Pour un langage régulier L sur Σ, on construit l'automate $\mathcal{A} = (Q, \Sigma, T, I, F)$ où

- $Q = \{q_{L'} \mid L'$ est un résiduel de $L\}$
- $I = \{q_L\}$
- $F = \{q_{L'} \mid \varepsilon \in L'\}$
- $(q_{L_1}, a, q_{L_2}) \in T$ ssi $L_2 = a^-1(L_1)$

Construire l'automate des résiduels de

- $a^* b^*$
- $aa(ba)^*$

Notez que cet automate est déterministe.

L'automate des résiduels est correct

Soit $w \in \Sigma^*$ et A l'automate des résiduels de L.

Montrons par récurrence sur $|w|$ que la lecture de w dans A termine dans l'état associé au résiduel $w^{-1}(L)$.

- si $|w| = 0$, alors $w = \varepsilon$, et le seul chemin étiquetté par ε est celui restant dans l'état initial: q_L et $L = \varepsilon^-1(L)$
- si $|w| = n > 0$, $w = ua$ où $u \in \Sigma^*$ et $a \in \Sigma$. Le chemin étiquetté par u finit dans l'état $q_{u^{-1}(L)}$ par hypothèse de récurrence. De plus, la seule transition étiquetté par a sortant de $q_{u^{-1}(L)}$ est $(q_{u^{-1}(L)}, a, q_{a^{-1}(u^{-1}(L))})$ et $a^{-1}(u^{-1=(L)}) = (ua)^{-1}(L) = w^{-1}(L)$ (voir DM)

Par ailleurs, un mot w est dans L ssi $\varepsilon \in w^{-1}(L)$.
Donc A accepte w ssi $w \in L$.

L'automate des résiduels est optimal

Soit L un langage, A l'automate des résiduels de L et A' un autre automate déterministe complet ayant moins d'états que A.

A' étant déterministe et complet, il existe une paire de mots w_1 et w_2 tels que $w_1^{-1}(L) \neq w_2^{-1}(L)$ et un état q de A' tel que les chemins étiquettés par w_1 et w_2 terminent en q.
$\rightarrow$ Il suffit de prendre un mot permettant d'accéder à chaque état de A, de voir dans quel état de A' ce mot termine, et d'appliquer le principe des tiroirs.

Soit $u \in w_1^{-1}(L) \setminus w_2^{-1}(L)$. On a donc que $w_1 u \in L$ et $w_2 u \notin L$. Hors, par déterminisme, A' accepte soit $w_1 u$ ET $w_2 u$ soit n'accepte ni $w_1 u$, ni $w_2 u$.

Donc A' ne reconnait pas L.

Alternative au lemme de l'étoile

Theorem (Théorème de MYHILL-NERODE)

Un langage est régulier si et seulement s'il possède un nombre fini de résiduels.

Alternative à l'automate des résiduels

$\longrightarrow$ Algorithme de création de l'automate minimale depuis une expression régulière.

La minimisation sans passer par le langage est possible : congruence de Nérode ou minimisation de Brzozowski par exemple.