Chapitre 2

Langages rationnels

Description

On peut décrire ces langages de plusieurs façons :

- par des grammaires régulières (ou linéaires);
- par des expressions régulières;
- par des automates d'états finis (A.E.F.);

Grammaires régulières

On dit qu'une grammaire G = (N, T, P, S) est régulière (ou linéaire) à gauche si toutes les règles de production sont de la forme :

Soient A, B $\in$ N et a $\in$ T*,

$$A \rightarrow Ba \quad \text{ou} \quad A \rightarrow a$$

On définit de manière duale les grammaires régulières à droite.

$S \rightarrow S10 \mid 0$

$S \rightarrow 0A$
$A \rightarrow 10A \mid \varepsilon$

Grammaires régulières

Théorème
Les grammaires régulières à gauche (ou à droite) engendrent les langages rationnels. A tout langage rationnel L correspond au moins une grammaire linéaire à droite et une grammaire linéaire à gauche.

Expressions régulières

Une expression régulière est définie inductivement :
Base: ∅, ε et les caractères de Σ sont des expressions régulières représentant les langages ∅, {ε} et {a} pour tout caractère «a» de Σ;
Règles: *si* r et s sont deux expressions régulières décrivant respectivement les langages R et S *alors* (r+s), (r.s) et r* sont des expressions régulières représentant respectivement les langages R∪S, R.S et R*

Exemples

Exemple d'expressions régulières sur Σ={a,b}:

- (a+b)* est l'ensemble des mots sur {a,b};
- (a+b)*.aa.(a+b)* est l'ensemble des mots contenant au moins une instance du motif « aa »;
- b+.aa.(a+b)* est l'ensemble des mots sur {a,b} qui commencent par un ou plusieurs « b », suivi du facteur « aa »;

Fermeture par les opérations régulières

Soient R et S deux langages sur Σ, alors :

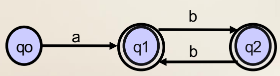
- Si R et S sont rationnels, alors R ∪ S est rationnel;
- Si R et S sont rationnels, alors RS est rationnel;
- Si R est rationnel, alors R* est rationnel.

Automates d'états finis (A.E.F.)

Un A.E.F. A est un quintuplet A = (E, Σ, δ, qo, F) :

- E est un ensemble fini d'états;
- Σ est un alphabet;
- δ est la fonction de transition;
- qo est l'unique état initial;
- F ⊆ E: est l'ensemble des états finaux.

Exemple :

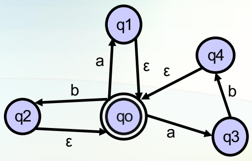


A.E.F. déterministes

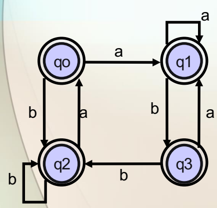
Un A.E.F. A est **déterministe** si à partir de tout état q_i il existe une unique transition pour chaque caractères de Σ.
Dans le cas contraire A est dit **non déterministe**;

Exemple

Exemple : L = { a, b, ab }*



Automate non déterministe



Automate déterministe

A.E.F. déterministes/non déterministes

Théorème :
Les A.E.F. déterministes, non déterministes et les A.E.F. avec ε-transitions reconnaissent les mêmes langages *: les langages rationnels*

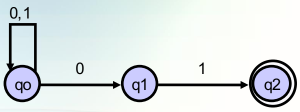
Fermeture par les opérations booléennes

Soient R et S deux langages sur Σ, alors :

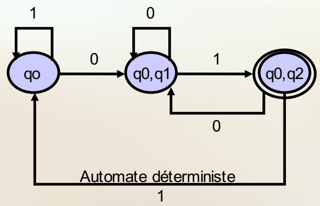
- Si R et S sont rationnels, alors $R \cup S$ est rationnel;
- Si R et S sont rationnels, alors $R \cap S$ est rationnel;
- Si R est rationnel, alors $\overline{R} = \Sigma^* - R$ est rationnel.

Exemple

Exemple : L = (0 + 1)* 01



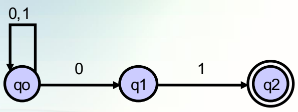
Automate non déterministe



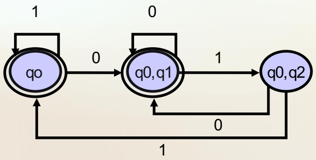
Automate déterministe

Exemple (suite)

Exemple : L = (0 + 1)* 01



Automate non déterministe



Automate déterministe complémentaire

Lois de Morgan

Les trois opérateurs booléens ne sont pas indépendants .

D'après les lois de DeMorgan nous avons :

- $R \cap S = \overline{R} \cup \overline{S}$ et $R \cup S = \overline{R} \cap \overline{S}$

Automate d'états finis minimum

Théorème
Soit L un langage rationnel. Il existe un A.E.F. déterministe minimum unique qui reconnaît le langage L.

Les langages rationnels sont donc non ambiguës.

Théorème de Kleene

Théorème
L'ensemble des langages rationnels sur un alphabet Σ est exactement l'ensemble des langages sur Σ reconnaissables par un automate d'état fini.

Stephen C. Kleene, 1956

Frontière

Théorème
Il existe des langages non rationnels.

Lemme de la Pompe (énoncé)

Lemme de la Pompe (Pumping Lemma)

Soit L un langage rationnel, il existe une constante n telle que pour tout mot w de L, avec $|w| \geq n$, il existe une décomposition de w en uvz avec :

$|uv| \leq n$ et $|v| \geq 1$ et telle que pour tout $i \geq 0$, $uv^iz \in L$

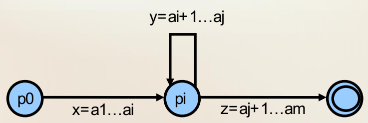
On utilise la contraposée du lemme de la Pompe pour montrer qu'un langage donné n'est pas rationnel.

Lemme de la Pompe (démonstration)

Si L est rationnel alors il existe un A.E.F. déterministe A tel que L=L(A). Soit n le nombre d'états de A.

Considérons tous les mots w de L(A) de taille m supérieure à n;
Définissons l'ensemble des états par lesquels passe A à la lecture de w :

$p_i = \delta(qo, a_1a_2 \ldots a_i)$, puisque $|w|>n$ il existe i et j tels que $p_i=p_j$;



Lemme de la Pompe (contraposée)

Contraposée du Lemme de la Pompe

Soit L un langage, si, quelque soit la constante n, il existe un mot w de L avec $|w| \geq n$ tel que pour toute décomposition de w en uvz avec $|uv| \leq n$ et $|v| \geq 1$ il existe $i \geq 0$ tel que uv^iz n'est pas dans L alors L n'est pas rationnel.

Exemple d'application

Montrons que le langage L des mots bien parenthésés n'est pas rationnel.

$L=\{w \in \{a, b\}^* \mid |w|_a = |w|_b$ et tout préfixe propre de w contient plus de 'a' que de 'b'$\}$;

Soit la constante n, considérons $w = a^{n+1} b^{n+1}$

On a bien $w \in L$;

Toute décomposition de w en uvz avec $|uv| \leq n$ impose que u et v ne sont composés que de 'a'.

Or pour, par exemple, i=2, le mot $uv^2 z$ n'est pas dans L car ce mot possède plus de 'a' que de 'b';

L n'est donc pas rationnel;

Utilisation de la pile

Pour vérifier qu'un mot est bien parenthésé, on doit utiliser une pile;

Pour chaque caractère rencontré, de nature ouvrante, on le stocke dans la pile;

S'il est de nature fermante, il doit fermer le caractère ouvrant de tête de pile;

On utilisera donc les automates à pile.

Pour résumer

- Les langages rationnels sont décrits par des expressions régulières, des A.E.F. ou des grammaires linéaires;
- Pour tout langage rationnel, il existe un unique A.E.F. minimum qui le reconnaît. Les langages rationnels sont donc non ambiguës;
- Il existe des langages non rationnels et on utilise le lemme de la Pompe pour démontrer facilement qu'ils ne sont pas rationnels.