Théorie des langages rationnels : THLR

CM 4

Uli Fahrenberg

EPITA Rennes

S3 2022

Aperçu

Programme du cours

1. Mots, langages
2. Langages rationnels, expressions rationnelles
3. Automates finis
4. Langages non-rationnels
5. Langages reconnaissables, minimisation

Dernièrement : Expressions rationnelles, langages rationnels

- poly chapitre 3, sections 3.1.1 et 3.1.2
- plus démonstration que L rationnel $\Rightarrow$ Pref(L) rationnel

Dernièrement : Expressions rationnelles

Soit Σ un alphabet.

Définition

Les expressions rationnelles sur Σ :

1. $\varnothing$ et ε sont des expressions rationnelles
2. pour tout $a \in \Sigma$, a est une expression rationnelle
3. e_1 et e_2 expressions rationnelles $\Rightarrow$ $e_1 + e_2$, $e_1.e_2$ et e_1^* aussi

Définition

Le langage dénoté par une expression rationnelle e sur Σ :

1. $L(\varnothing) = \emptyset$, $L(\varepsilon) = \{\varepsilon\}$
2. $L(a) = \{a\}$ pour tout $a \in \Sigma$
3. $L(e_1 + e_2) = L(e_1) \cup L(e_2)$, $L(e_1.e_2) = L(e_1).L(e_2)$,
   $L(e^*) = (L(e))^*$

Dernièrement : Langages rationnels

Définition

Les langages rationnels sur Σ :

1. $\emptyset$ et $\{\varepsilon\}$ sont des langages rationnels
2. pour tout $a \in \Sigma$, $\{a\}$ est un langage rationnel
3. L_1 et L_2 langages rationnels $\Rightarrow$ $L_1 \cup L_2$, $L_1.L_2$ et L_1^* aussi

Théorème

$L \subseteq \Sigma^*$ est rationnel ssi il existe une expression rationnelle e telle que $L = L(e)$.

5 minutes de réflexion

Vrai ou faux ?

1. Si L_1 et L_2 sont rationnels, alors $L_1 \cup L_2$ est rationnel
2. Si L_1 et L_2 sont rationnels, alors $L_1 \cap L_2$ est rationnel
3. Chaque sous-ensemble d'un langage rationnel L est rationnel.

Pour chaque expression rationnelle suivante sur alphabet $\Sigma = \{a, b\}$, trouvez deux mots qui appartiennent de leur langage et deux autres qui ne l'appartiennent pas :

4. $a^* b^*$
5. $a^* + b^*$
6. $(aaa)^*$
7. $(a + b)^* ab\,(a + b)^* ba\,(a + b)^*$
8. $(a^* b)^*(b^* a)^*$

5 minutes de réflexion

Vrai ou faux ?

1. Si L_1 et L_2 sont rationnels, alors $L_1 \cup L_2$ est rationnel ✓
2. Si L_1 et L_2 sont rationnels, alors $L_1 \cap L_2$ est rationnel ✓
3. Chaque sous-ensemble d'un langage rationnel L est rationnel. ✗

Pour chaque expression rationnelle suivante sur alphabet $\Sigma = \{a, b\}$, trouvez deux mots qui appartiennent de leur langage et deux autres qui ne l'appartiennent pas :

4. $a^* b^*$
5. $a^* + b^*$
6. $(aaa)^*$
7. $(a+b)^* ab\,(a+b)^* ba\,(a+b)^*$
8. $(a^* b)^* (b^* a)^*$

Bonus

Bonus : monoïdes et demi-anneaux

La structure $(\Sigma^*, ., \varepsilon)$ des mots sur Σ forme un monoïde.

- comme un groupe, mais sans inverses
- (et pas commutative)

En fait, le monoïde libre sur Σ.

- donc tout monoïde est un quotient d'un monoïde Σ^* pour quelque Σ

La structure $(\mathcal{P}(\Sigma^*), \cup, ., \emptyset, \{\varepsilon\})$ des langages sur Σ forme un demi-anneau.

- comme un anneau, mais sans inverses additifs
- langages finis sur Σ : le demi-anneau idempotent libre sur Σ

Avec l'étoile de Kleene, $(\mathcal{P}(\Sigma^*), \cup, ., ^*, \emptyset, \{\varepsilon\})$ forme un algèbre de Kleene.

- structure algébrique fondamentale pour l'informatique
- mais c'est quoi les algèbres de Kleene libres ?

Bonus : algèbres de Kleene

Un demi-anneau est une structure algébrique $(S, \oplus, \otimes, \mathbb{0}, \mathbb{1})$ telle que

- $(S, \oplus, \mathbb{0})$ forme un monoïde commutatif,
- $(S, \otimes, \mathbb{1})$ forme un monoïde,
- $x(y \oplus z) = xy \oplus xz$, $(x \oplus y)z = xz \oplus yz$ et $x\mathbb{0} = \mathbb{0}x = \mathbb{0}$

S est idempotent si $x \oplus x = x$.

Théorème

L'ensemble de langages finis forme le demi-anneau idempotent libre.

Une algèbre de Kleene est un demi-anneau idempotent S équipé avec toutes les sommes géométriques $\bigoplus_{n\geq 0} x^n$, pour tout $x \in S$, et telle que $x \otimes \left(\bigoplus_{n\geq 0} y^n\right) \otimes z = \bigoplus_{n\geq 0} (xy^n z)$ pour tout $x, y, z \in S$.

Théorème

L'ensemble de langages rationnels forme l'algèbre de Kleene libre.

Un peu de maths

Nombres

- des entiers naturels : $\mathbb{N} = \{0, 1, 2, \dots\}$
- des entiers relatifs : $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$
- des nombres rationnels : $\mathbb{Q} = \{\frac{a}{b} \mid a \in \mathbb{Z}, b \in \mathbb{Z}, b \neq 0\}$
- des nombres réels : $\mathbb{R} = ?$
- (des nombres complexes : *on s'en fout ici*)

Construction

Die natürlichen Zahlen hat Gott gemacht, alles andere ist Menschenwerk

— L. Kronecker 1886

- de $\mathbb{N}$ à $\mathbb{Z}$: $\mathbb{N} \times \mathbb{N}$ modulo la relation d'équivalence
$$(x_1, y_1) \sim (x_2, y_2) \iff x_1 + y_2 = x_2 + y_1$$
- de $\mathbb{Z}$ à $\mathbb{Q}$: $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ modulo la relation d'équivalence
$$(x_1, y_1) \sim (x_2, y_2) \iff x_1 y_2 = x_2 y_1$$
- de $\mathbb{Q}$ à $\mathbb{R}$: via des suites convergentes / suites de Cauchy :
  - soit $S = \{(x_0, x_1, \dots) \in \mathbb{Q}^\infty \mid \lim_{m,n \to \infty} (x_m - x_n) = 0\}$
  - soit $\sim$ la relation d'équivalence sur S défini par
$$(x_0, x_1, \dots) \sim (y_0, y_1, \dots) \iff \lim_{m,n \to \infty} (x_m - y_n) = 0$$
  - alors $\mathbb{R} = S_{/\sim}$

Dénombrabilité

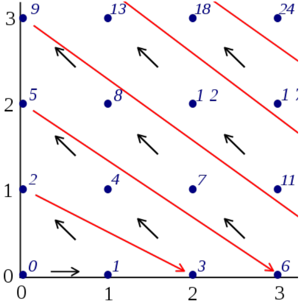
Définition

Un ensemble S est dénombrable s'il existe une bijection $f : \mathbb{N} \to S$.

- $\mathbb{N}$ est triviellement dénombrable.
- $\mathbb{Z}$ est dénombrable via la bijection $f(n) = \begin{cases} -\frac{n}{2} & \text{si } n \text{ est pair,} \\ \frac{n+1}{2} & \text{si } n \text{ est impair :} \end{cases}$

$$\mathbb{Z} = \{0, 1, -1, 2, -2, \dots\}$$

- $\mathbb{Q}^+$ est dénombrable comme suite :



Argument de la diagonale de Cantor

Théorème (G. Cantor 1891)

$\mathbb{R}$ *n'est pas dénombrable.*

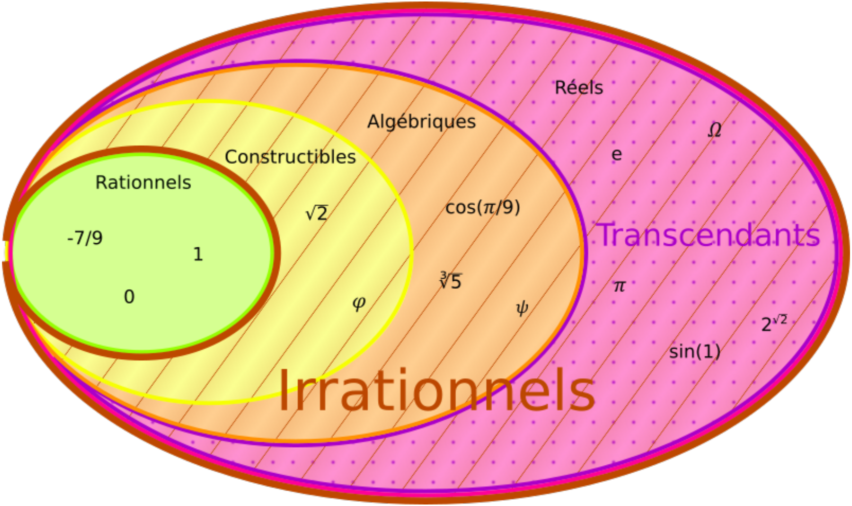
Démonstration.

1. Supposons que $\mathbb{R}$ soit dénombrable, alors l'intervalle ouvert $S = \{x \mid 0 < x < 1\}$ l'est aussi.
2. Soit $E = \{x_0, x_1, \dots\}$ une énumération de S. Notons alors

$$\begin{aligned} x_0 &= 0{,}c_{00}\, c_{01}\, c_{02} \dots \\ x_1 &= 0{,}c_{10}\, c_{11}\, c_{12} \dots \\ x_2 &= 0{,}c_{20}\, c_{21}\, c_{22} \dots \\ &\vdots \end{aligned}$$

3. Soit $d_n = 9 - c_{nn}$ pour tout $n \geq 0$ et $y = 0{,}d_0 d_1 d_2 \dots$
4. Alors $y \in S$, mais $y \neq x_n$ pour tout $n \geq 0$, donc $y \notin E$. $\lightning$

Nombres réels



Bonus bonus

Vous vous souvenez ?

Définition

Un langage L est récursivement énumerable s'il existe un algorithme qui énumère tout les mots de L.

Théorème

Il existe un langage qui n'est pas récursivement énumerable.

Démonstration.

1. L'ensemble de tous algorithmes est dénombrable. *(Pourquoi ? Qu'est-ce que ?)*
2. Chaque algorithme n'énumère guère qu'un langage.
3. L'ensemble de langages n'est pas dénombrable. *(Pourquoi ?)*

L'ensemble de langages n'est pas dénombrable

- Soit Σ un alphabet (un ensemble fini non-vide)
- Un langage est un sous-ensemble $L \subseteq \Sigma^*$

$\Rightarrow$ L'ensemble de langages : $\mathcal{P}(\Sigma^*)$

Théorème

$\mathcal{P}(\Sigma^*)$ *n'est pas dénombrable.*

Démonstration.

L'ensemble de langages n'est pas dénombrable

- Soit Σ un alphabet (un ensemble fini non-vide)
- Un langage est un sous-ensemble $L \subseteq \Sigma^*$

$\Rightarrow$ L'ensemble de langages : $\mathcal{P}(\Sigma^*)$

Théorème

$\mathcal{P}(\Sigma^*)$ *n'est pas dénombrable.*

Démonstration.

1. Supposons que $\mathcal{P}(\Sigma^*)$ soit dénombrable, alors le sous-sensemble $\mathcal{J} = \{L \subseteq \Sigma^* \mid L \text{ infini } \}$ l'est aussi.
2. Soit $E = \{L_0, L_1, \dots\}$ une énumération de $\mathcal{J}$. Chaque L_i est dénombrable, alors notons $L_i = \{w_{i,0}, w_{i,1}, \dots\}$.
3. Soit $L = \bigcup_{i=0}^{\infty}(L_i \setminus \{w_{i,i}\})$.
4. Alors $L \in \mathcal{J}$, mais $L \neq L_i$ pour chaque i : il n'est pas dans notre énumération E.

$\lightning$


That's all Folks!