Théorie des langages rationnels : THLR CM 2

Uli Fahrenberg

EPITA Rennes

Septembre 2021



Aperçu

Programme du cours

Aperçu

000000

- Mots, langages
- Langages rationnels, expressions rationnelles
- Automates finis
- Langages non-rationnels
- Langages reconnaissables, minimisation

Hier : L'algèbre de mots

Soit Σ un ensemble fini.

• on appelle les éléments $a, b, \ldots \in \Sigma$ des symboles

On dénote Σ^* l'ensemble de tous les suites finies d'éléments de Σ .

- donc $\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \cdots = \bigcup_{n \geq 0} \Sigma^n$
- on appelle les éléments $u, v, w, \ldots \in \Sigma^*$ des mots

La concaténation de deux mots $a_1 \dots a_n$ et $b_1 \dots b_m$ est le mot

$$a_1 \ldots a_n b_1 \ldots b_m$$

- E: le mot vide
- l'opération « . » sur mots est associative et a ε comme

élément neutre de deux côtés

La longueur |u| d'un mot $u \in \Sigma^*$: le nombre de symboles de u.

- $|\varepsilon| = 0$ et |uv| = |u| + |v|
- u^n : la concaténation de n copies de u
- $\bullet |u^n| = n|u|$

Hier : L'algèbre de langages

Un langage est un sous-ensemble $L \subseteq \Sigma^*$.

- opérations ensemblistes : $L_1 \cup L_2$, $L_1 \cap L_2$, \overline{L}
- concaténation : $L_1 L_2 = \{u_1 u_2 \mid u_1 \in L_1, u_2 \in L_2\}$
- $L^n = L \cdots L$ (n copies de L)
- étoile de Kleene : $L^* = L^0 \cup L_1 \cup L^2 \cup \cdots = \bigcup_{n \ge 0} L^n$

L'opération « . » sur langages est associative et a $\{\varepsilon\}$ comme élément neutre de deux côtés.

- $L.\emptyset = \emptyset.L = \emptyset$
- $\emptyset^* = \{\varepsilon\}^* = \{\varepsilon\}$

Pour aller plus loin

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?

5 minutes de réflexion

Vrai ou faux?

- $\{a\}^n = \{a^n\}$
- $\{a,b\}^n = \{a^n,b^n\}$
- \bullet L^* est un ensemble infini pour tout $L \subseteq \Sigma^*$
- pour tout $L \subseteq \Sigma^*$ et $n \in \mathbb{N}$, l'ensemble $\{u \in L \mid |u| \leq n\}$ est un ensemble fini

- $(L_1 \cup L_2)^* = (L_1^* L_2)^* L_1^*$

5 minutes de réflexion

Vrai ou faux?

$$\{a\}^n = \{a^n\}$$

$$\{a,b\}^n = \{a^n,b^n\}$$

$$ullet$$
 L^* est un ensemble infini pour tout $L\subseteq \Sigma^*$

• pour tout
$$L \subseteq \Sigma^*$$
 et $n \in \mathbb{N}$, l'ensemble $\{u \in L \mid |u| \leq n\}$ est un ensemble fini

$$(L_1 \cup L_2)^* = (L_1^* L_2)^* L_1^*$$

Uli Fahrenberg

Langages rationnels

Opérations rationnelles

Soit Σ un alphabet, on travaille avec des langages dans $\mathcal{P}(\Sigma^*)$.

On a vu dans DM1-exo1 que les opérations \cup , . et * sont bien spéciales.

- $Pref({a}{b}^*) = {\varepsilon, a} \cup {a}{b}^*$
- Suff($\{a\}\{b\}^*$) = ???

Définition

Les opérations rationnelles dans $\mathcal{P}(\Sigma^*)$ sont \cup , . et *.

• donc union, concaténation et étoile de Kleene

Théorème (pour plus tard) : Toutes les autres opérations sont exprimables par \cup , . et *.

Définition (3.1)

Les langages rationnels sur Σ sont définis inductivement comme suite :

- \bigcirc \emptyset et $\{\varepsilon\}$ sont des langages rationnels
- ② pour tout $a \in \Sigma$, $\{a\}$ est un langage rationnel
- \odot si L_1 et L_2 sont des langages rationnels, alors $L_1 \cup L_2$, $L_1.L_2$ et L_1^* le sont également
 - $\{\varepsilon\} = \emptyset^* \Rightarrow$ on peut enlever $\{\varepsilon\}$ de la définition

Lemme

L est rationnel si et seulement si

- $L = \emptyset$ ou $L = \{a\}$ pour un $a \in \Sigma$ ou
- $L = L_1 \cup L_2$, $L = L_1 L_2$ ou $L = L_1^*$ pour L_1 et L_2 rationnels.

(En quoi ce lemme est-il différent de la définition?)

Rationalité

Théorème

Si L_1 et L_2 sont des langages rationnels, alors $L_1 \cap L_2$, \overline{L}_1 , $\mathsf{Pref}(L_1)$, $\mathsf{Suff}(L_1)$ et $\mathsf{Fact}(L_1)$ le sont aussi.

• pour la démonstration faut attendre quelques journées

5 minutes de réflexion

Rationnel ou pas rationnel, sur alphabet $\Sigma = \{a, b, c\}$?

- § a, b, abcba
- ② $\{a^n \mid n > 0\}$
- ∫ {*w* ∈ Σ* | |*w*| ≥ 5}
- $\{a^{2n} \mid n \geq 0\}$
- $\{a^{n^2} \mid n \geq 0\}$
- $\{a^m b^n \mid m, n \ge 0\}$
- $\{a^n b^n \mid n \ge 0\}$

5 minutes de réflexion

Rationnel ou pas rationnel, sur alphabet $\Sigma = \{a, b, c\}$?

$$\{a^n \mid n > 0\}$$

$$\{ w \in \Sigma^* \mid |w| \ge 5 \}$$

$$\{a^{2n} \mid n \geq 0\}$$

$$\{a^{n^2} \mid n \ge 0\}$$

$$\{a^m b^n \mid m, n \ge 0\}$$

$$\{a^nb^n \mid n > 0\}$$

Une notation pratique pour des langages rationnels :

Définition (3.2)

Les expressions rationnelles sur Σ sont définis inductivement comme suite :

- \bigcirc Ø et ε sont des expressions rationnelles
- ② pour tout $a \in \Sigma$, a est une expression rationnelle
- \odot si e_1 et e_2 sont des expressions rationnelles, alors e_1+e_2 , $e_1.e_2$ et e_1^* le sont également

Une notation pratique pour des langages rationnels :

Définition (3.1, recall)

Les langages rationnels sur Σ sont définis inductivement comme suite :

- **1** \emptyset et $\{\varepsilon\}$ sont des langages rationnels
- ② pour tout $a \in \Sigma$, $\{a\}$ est un langage rationnel
- ullet si L_1 et L_2 sont des langages rationnels , alors $L_1 \cup L_2$, $L_1.L_2$ et L_1^* le sont également
 - presque la même chose! mais
- 3.1 introduit une classe de sous-ensembles de Σ^* ,
- 3.2 définit des expressions syntaxiques

Une notation pratique pour des langages rationnels :

Définition (3.2)

Les expressions rationnelles sur Σ sont définis inductivement comme suite :

- \bigcirc et ε sont des expressions rationnelles
- ② pour tout $a \in \Sigma$, a est une expression rationnelle
- \odot si e_1 et e_2 sont des expressions rationnelles, alors e_1+e_2 , $e_1.e_2$ et e_1^* le sont également
 - presque la même chose! mais
 - 3.1 introduit une classe de sous-ensembles de Σ^* ,
 - 3.2 définit des expressions syntaxiques

On va relier les deux en donnant une sémantique aux expressions rationnelles.

Sémantique

Définition

Le langage dénoté par une expression rationnelle e sur Σ est $L(e) \subseteq \Sigma^*$ définit inductivement comme suite :

- ② $L(a) = \{a\}$ pour tout $a \in \Sigma$
- $L(e_1 + e_2) = L(e_1) \cup L(e_2), \ L(e_1.e_2) = L(e_1).L(e_2), \ L(e^*) = (L(e))^*$

Théorème

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

Démonstration.

Par induction structurelle (sur tableau).

La démonstration (sur tableau)

Les langages rationnels sur Σ :

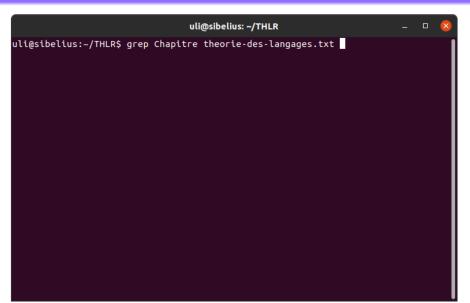
- **1** \emptyset et $\{\varepsilon\}$ sont des langages rationnels
- ② pour tout $a \in \Sigma$, $\{a\}$ est un langage rationnel
- **③** L_1 et L_2 languages rationnels ⇒ $L_1 \cup L_2$, $L_1.L_2$ et L_1^* aussi

Les expressions rationnelles sur Σ :

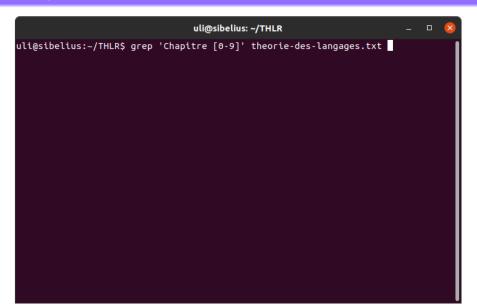
- \bigcirc et ε sont des expressions rationnelles
- ② pour tout $a \in \Sigma$, a est une expression rationnelle
- \bigcirc e_1 et e_2 expressions rationnelles \Rightarrow e_1+e_2 , $e_1.e_2$ et e_1^* aussi

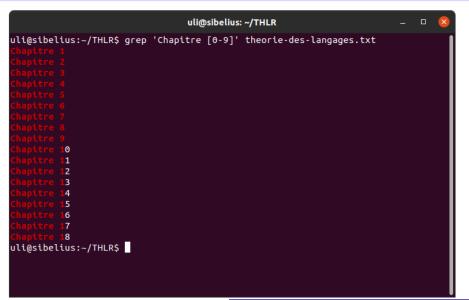
Le langage dénoté par e :

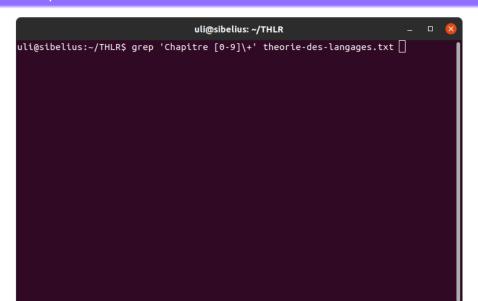
- ② $L(a) = \{a\}$ pour tout $a \in \Sigma$
- ① $L(e_1 + e_2) = L(e_1) \cup L(e_2), L(e_1.e_2) = L(e_1).L(e_2),$ $L(e^*) = (L(e))^*$

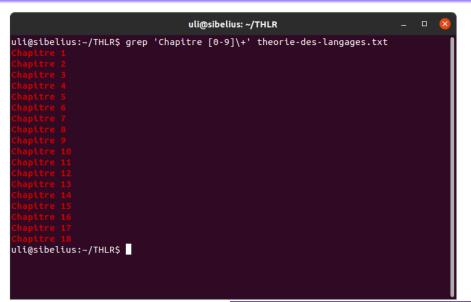


```
uli@sibelius: ~/THLR
uli@sibelius:~/THLR$ grep Chapitre theorie-des-langages.txt
         8
  apitre 10
        rédigé par Pierre Senellart.
        11
        rédigé par Pierre Senellart.
        12
         13
        14
         15
         16
         18
uli@sibelius:~/THLR$
```











```
uli@sibelius: ~/THLR
Dox
absolue. Vuibert. Bande dessiné
   du
Hodges, A. (1992). Alan Turing. The Enigma. Random House. Biographie d'Alan Turi
ng,
traduit en français sous le nom de Alan Turing ou l'énigme de l'intelligence. 11
Hopcroft, J. E. et Ullman, J. D. (1979). Introduction to automata theory, langua
Perr
14(4). Un historique de l'émergence de la théorie des automates. 11.2
Sakarovitch, J. (2003). Eléments de théorie des automates. Vuibert, Paris. 4
Turing, A. M.
                Proceedings of the London Mathematical Society, 2(42), L'article i
ntroduisant
la notion de machine de Turin
Intéressant à la fois par son contenu et par son aspect historique. 11.2
expression rationnel

    lextcographique

                    , 17
transformat<mark>ion syntaxique,</mark> 27 uli@sibelius:~/THLR$
```



uli@sibelius: ~/THLR st connu pour les automates cellulaires, le tri fusion, et bien sûr prestigieuse médaille John von Neumann à un chercheur en informatiqu est un informaticien notable pour ses contributions à la fois théoriques et est souvent considéré comme le premier informaticien. Ses travaux sur le plus haute distinction que peut recevoir un chercheur en informatique. Automate 12.11 – Le dictionnaire déterminisé On conçoit aisément que pour un dictionnaire du français, ce principe de « factorisa tion » Automate 12.12 – Le dictionnaire déterminisé et minimisé . . . 168 . . 169 Barsky, R. F. (1997). Noam Chomsky: A Life of Dissent. MIT Press. B<mark>iographie</mark> de Noam Hodges, A. (1992). Alan Turing. The Enigma. Random House. B<mark>iographie</mark> d'Alan Turing, Perrin, D. (1995). Les débuts de la théorie des automates. Technique et science 🕆 la notion de machine de Turing, un des articles fondateurs de la science informatiq lexicographique, 17 uli@sibelius:~/THLRS

Théorème

Si L est un langage rationnel, alors Pref(L) l'est aussi.

Théorème

Si L est un langage rationnel, alors Pref(L) l'est aussi.

Démonstration.

① Soit e une expression rationnelle telle que L = L(e).

Théorème

Si L est un langage rationnel, alors Pref(L) l'est aussi.

- ① Soit e une expression rationnelle telle que L = L(e).
- Nous construirons une expression rationnelle pref(e) telle que L(pref(e)) = Pref(L).

Théorème

Si L est un langage rationnel, alors Pref(L) l'est aussi.

- Soit e une expression rationnelle telle que L = L(e).
- Nous construirons une expression rationnelle pref(e) telle que $L(\operatorname{pref}(e)) = \operatorname{Pref}(L)$
- ... par induction structurelle :

Théorème

Si L est un langage rationnel, alors Pref(L) l'est aussi.

- ① Soit e une expression rationnelle telle que L = L(e).
- Nous construirons une expression rationnelle pref(e) telle que L(pref(e)) = Pref(L).
- ... par induction structurelle :

Théorème

Si L est un langage rationnel, alors Pref(L) l'est aussi.

- ① Soit e une expression rationnelle telle que L = L(e).
- ② Nous construirons une expression rationnelle pref(e) telle que L(pref(e)) = Pref(L).
- ... par induction structurelle :
- $\operatorname{pref}(\varnothing) = \varnothing$, $\operatorname{pref}(\varepsilon) = \varepsilon$, $\operatorname{pref}(a) = a + \varepsilon$

Théorème

Si L est un langage rationnel, alors Pref(L) l'est aussi.

- ① Soit e une expression rationnelle telle que L = L(e).
- Nous construirons une expression rationnelle pref(e) telle que L(pref(e)) = Pref(L).
- ... par induction structurelle :
- $\operatorname{pref}(\varnothing) = \varnothing$, $\operatorname{pref}(\varepsilon) = \varepsilon$, $\operatorname{pref}(a) = a + \varepsilon$
- o pref $(e_1 + e_2) =$ pref $(e_1 e_2) =$ pref $(e^*) =$

Théorème

Si L est un langage rationnel, alors Pref(L) l'est aussi.

- ① Soit e une expression rationnelle telle que L = L(e).
- ② Nous construirons une expression rationnelle pref(e) telle que L(pref(e)) = Pref(L).
- ... par induction structurelle :
- $\operatorname{pref}(\varnothing) = \varnothing$, $\operatorname{pref}(\varepsilon) = \varepsilon$, $\operatorname{pref}(a) = a + \varepsilon$

Théorème

Si L est un langage rationnel, alors Pref(L) l'est aussi.

- ① Soit e une expression rationnelle telle que L = L(e).
- ② Nous construirons une expression rationnelle pref(e) telle que L(pref(e)) = Pref(L).
- ... par induction structurelle :
- $\operatorname{pref}(\varnothing) = \varnothing$, $\operatorname{pref}(\varepsilon) = \varepsilon$, $\operatorname{pref}(a) = a + \varepsilon$

Théorème

Si L est un langage rationnel, alors Pref(L) l'est aussi.

- ① Soit e une expression rationnelle telle que L = L(e).
- ② Nous construirons une expression rationnelle pref(e) telle que L(pref(e)) = Pref(L).
- Opening the structure of the structur
- $\operatorname{pref}(\varnothing) = \varnothing$, $\operatorname{pref}(\varepsilon) = \varepsilon$, $\operatorname{pref}(a) = a + \varepsilon$

Théorème

Si L est un langage rationnel, alors Pref(L) l'est aussi.

- ① Soit e une expression rationnelle telle que L = L(e).
- Nous construirons une expression rationnelle pref(e) telle que L(pref(e)) = Pref(L).
- Opening in the structure of the struc
- $\operatorname{pref}(\varnothing) = \varnothing$, $\operatorname{pref}(\varepsilon) = \varepsilon$, $\operatorname{pref}(a) = a + \varepsilon$
- **o** Maintenant il faut démontrer que, en fait, L(pref(e)) = Pref(L).
- ... par induction structurelle, encore (sur tableau).

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 : $\mathbb{Q}=\{\frac{a}{b}\mid a\in\mathbb{Z},b\in\mathbb{Z},b\neq0\}$ • des nombres réels : $\mathbb{R}=?$ • (des nombres complexes : on s'en fout)

Construction

Die natürlichen Zahlen hat Gott gemacht, alles andere ist Menschenwerk
— L. Kronecker 1886

 \bullet 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 $\mathbb Q:\mathbb Z\times \big(\mathbb Z\setminus\{0\}\big)$ modulo la relation d'équivalence

$$(x_1,y_1)\sim(x_2,y_2)\Longleftrightarrow x_1y_2=x_2y_1$$

- ullet 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
 - $(x_0,x_1,\dots)\sim (y_0,y_1,\dots)\Longleftrightarrow \lim_{m,n\to\infty}(x_m-y_n)=0$
 - alors $\mathbb{R} = S_{/\sim}$

Dénombrabilité

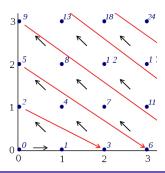
Définition

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

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

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

Q⁺ est dénombrable comme suite :



Argument de la diagonale de Cantor

Théorème (G. Cantor 1891)

R n'est pas dénombrable.

- 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

$$x_0 = 0, c_{00} c_{01} c_{02} \dots$$

 $x_1 = 0, c_{10} c_{11} c_{12} \dots$
 $x_2 = 0, c_{20} c_{21} c_{22} \dots$
:

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

Nombres réels

