Linguistique computationnelle

Chapitre 3

Expressions régulières

Table des matières

UNIVERSITÉ SORBONNE NOUVELLE PARIS 3
11 mars 2021

Déf. 6 (Opérations rationnelles)

On peut définir deux opérations binaires et une opération unaire sur les langages :

- L'*union* des langages est définie comme d'habitude (union ensembliste)
- Le *produit* des langages est défini de la manière suivante (on suppose l'opération de concaténation définie) : $L_1.L_2 = \{uv \,/\, u \in L_1 \text{ et } v \in L_2\}$
  En généralisant, on peut proposer la notation

$$\left(\begin{array}{lcl} A^0 & = & \{\varepsilon\} \\ A^1 & = & A \\ A^{i+1} & = & A.A^i \\ \vdots & & \\ A^n & = & \{a_1...a_n \,/\, a_i \in A\} \end{array}\right.$$

- L'*étoile* (ou fermeture) d'un langage est définie ainsi : $A^* = \bigcup_{n\geqslant 0} A^n$

$$\{0,1,2\}^* = \{\varepsilon\} \cup \{0,1,2\} \cup \{0,1,2\}\cdot\{0,1,2\} \cup \{0,1,2\}\cdot\{0,1,2\}\cdot\{0,1,2\} \cup \ldots$$

$$\underbrace{\{0,1,2\}}_{}^{3}\cdot\underbrace{\{0,1,2\}}_{}^{3} = \underbrace{\{00,01,02,10,11,12,20,21,22\}}_{}^{3\times 3=9}$$

$$\{0,1,2\}\cdot\{0,1,2\}\cdot\{0,1,2\} = \quad \cdot\{0,1,2\}$$

$$= \{000, 001, \ldots \quad 222\}$$

3.3 Automates et expressions rationnelles

3.3.1 Expressions rationnelles

Les opérations rationnelles peuvent être utilisées pour décrire des langages quelconques :

- $\{au, le, par, un\}.\{\_\}.\{matin, soir\}$ $= L_1$ $(\Sigma = \{a, b, c \ldots, z\})$
- $\{1, 2, \ldots, 9\}.\{0, 1, 2, \ldots, 9\}.\left(\{,\}.\{0, 1, 2, \ldots, 9\}^* \cup \{\varepsilon\}\right)$ L_2 $(\Sigma = \{0, 1, 2 \ldots, 9, ,\})$
- $\{aa\}.(\{b\}.\{b\} \cup \{cc\})^*$ $(\Sigma = \{a, b, c\})$
- $\{aa\}.(\{b\}^* \cup \{cc\})$ $= L_4$ $(\Sigma = \{a, b, c\})$

$\varepsilon \in$

$L_4 = \{aa, aab, aabb \ldots aacc\}$

$L_1 = \{$au␣matin, le␣soir ... le␣matin, par␣soir, par␣matin, un␣soir, un␣matin, au␣soir$\}$

$L_2 = \{\cancel{2,1};\ 20,13527;\ 97;$

$\{,\}.\{0,1,\ldots 9\}^* = \{,\ ;\ ,3\ ;\ ,7\ ;\ ,12345\ ;\ ,2222$ etc $\}$

$\{1,2 \ldots 9\}.\{0,1 \ldots 9\} = \{10, 11, 19, 99, 90 \ldots\}$

En particulier, on peut construire des expressions en n'utilisant que des langages choisis, qu'on peut appeler **la base** :

- $\Big(\{0\} \cup \{1\} \cup \ldots \{9\}\Big) . \Big(\{0\} \cup \{1\} \cup \ldots \{9\}\Big)^*$
- $\{a\}.\{a\}.\underbrace{(\{b\}^* \cup \{c\}.\{c\})}_{\{\varepsilon, b, bb, bbb \ldots\} \cup \{cc\}}$

$\{aa, b\} . \{cc, abc\}^*$

$\{9, 90, 987 \ldots, 07$

$\{0\} \cup \{1\} \cup \{2\} = \{0, 1, 2\}$

langages singleton (avec une seule lettre.

exclut

aabcc

bac

$\{aabb, aacc, aab \ldots$

aa

Pour $X = \{a, b, c, d\}$, On peut définir $X^* = \Big(\{a\} \cup \{b\} \cup \{c\} \cup \{d\}\Big)^*$

X^* : tous les mots qu'on peut former sur l'alphabet X

Notation simplifiée :

- $\{a\}$ sera simplement noté a
- $\{\varepsilon\}$ sera simplement noté ε
- la réunion $\cup$ se notée $|$
- le point de la concaténation sera ignoré

$\underline{\{a\}}$ $\underline{\{b\}}$

$$(b|aa)^*ab^* = \underbrace{(\{b\} \cup \{a\}.\{a\})^* .\{a\}.\{b\}^*}$$

$ab^* \neq (ab)^*$

$2 \times 2 \geq 4$

$\{a,b\} . \{b,c\} = \{ab, ac, bb, bc\}$

$\{a\} . \{b\} = \{ab\}$

$3 + 2 \geq$

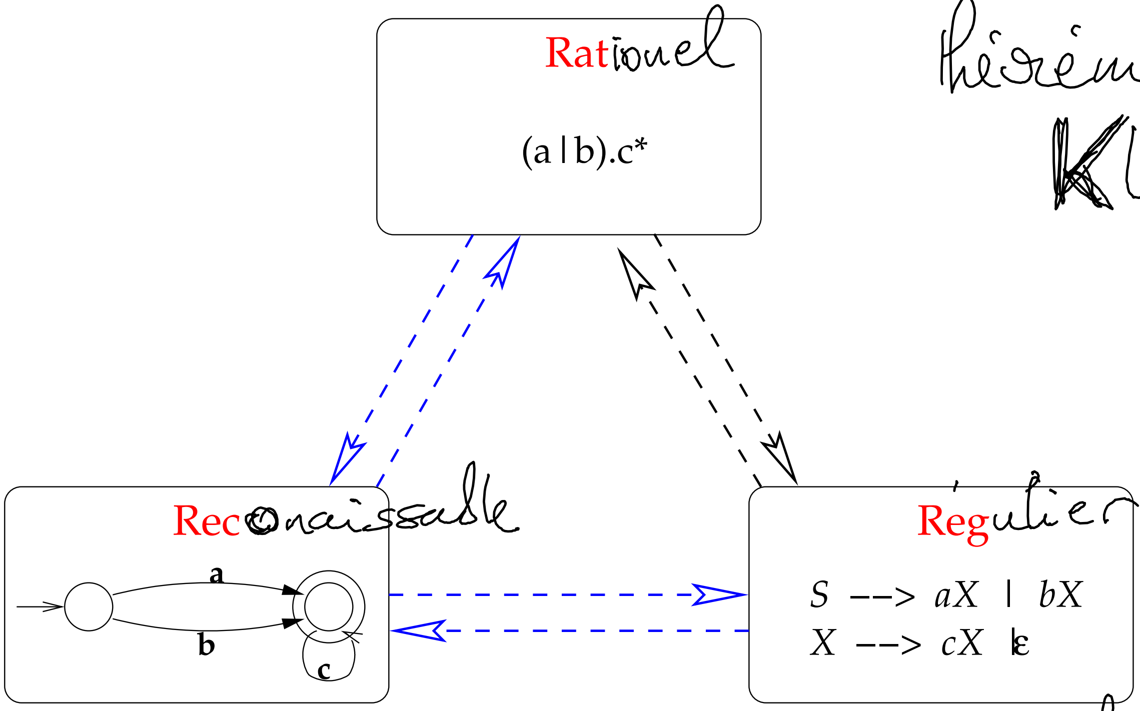
$\{a,b,c\} \cup \{e,f\} = \{a,b,c,e,f\}$

$\{a\} \cup \{b\} = \{a,b\}$

3.3.2 Correspondance avec les automates

Résultat principal

Théorème de Kleene

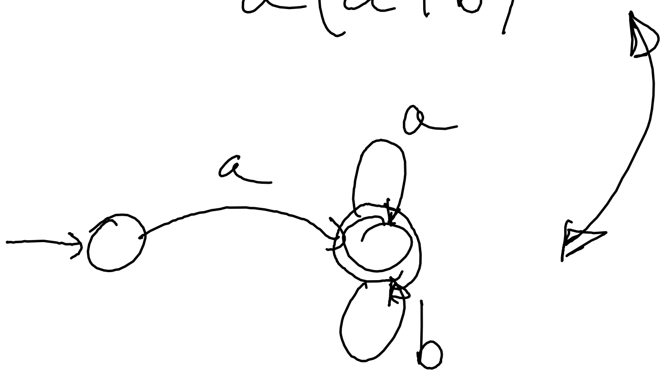


(grammaires formelles)

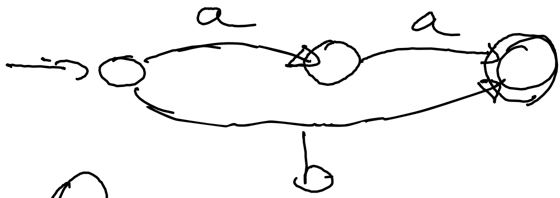
Correspondance : exemples

(1)
- a. $(a|b)^*$
- b. $a(a|b)^*$
- c. $(b|ab)^*(a|\varepsilon)$
- d. $a^*|b^*$
- e. $(aa|b)^*$
- f. $(ab^*a|b)^*$

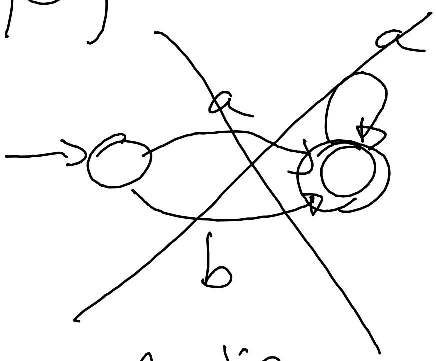
$a(a|b)^*$

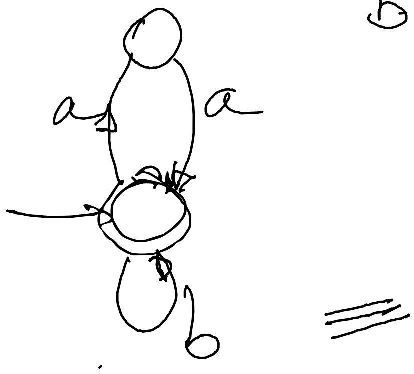


$(aa|b)$



$(aa|b)^*$

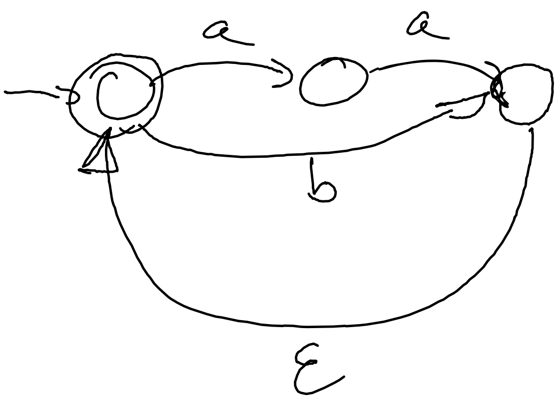




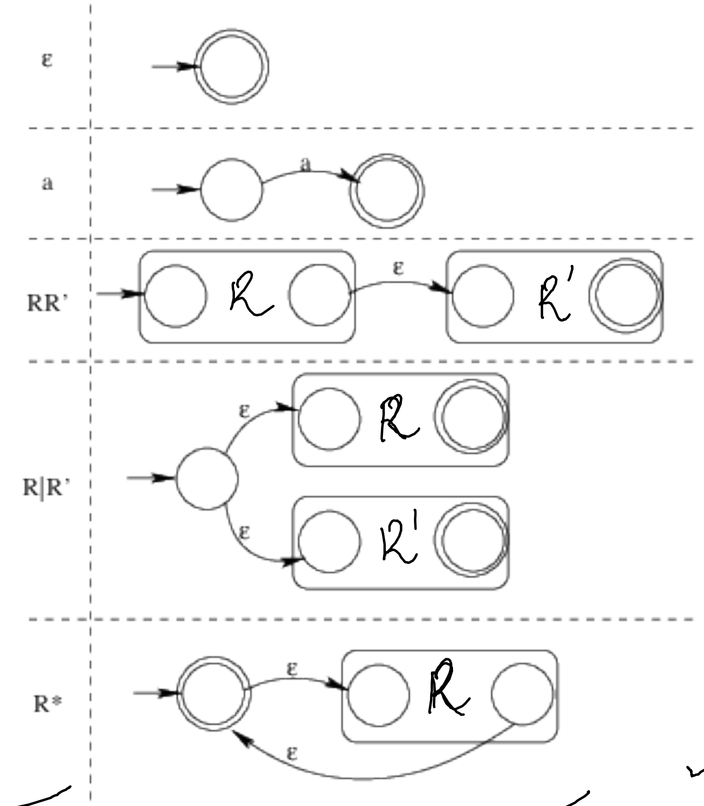
$ba \notin \mathcal{L}$

$aaa \notin \mathcal{L}$

$bbb \in \mathcal{L}$



Principe général

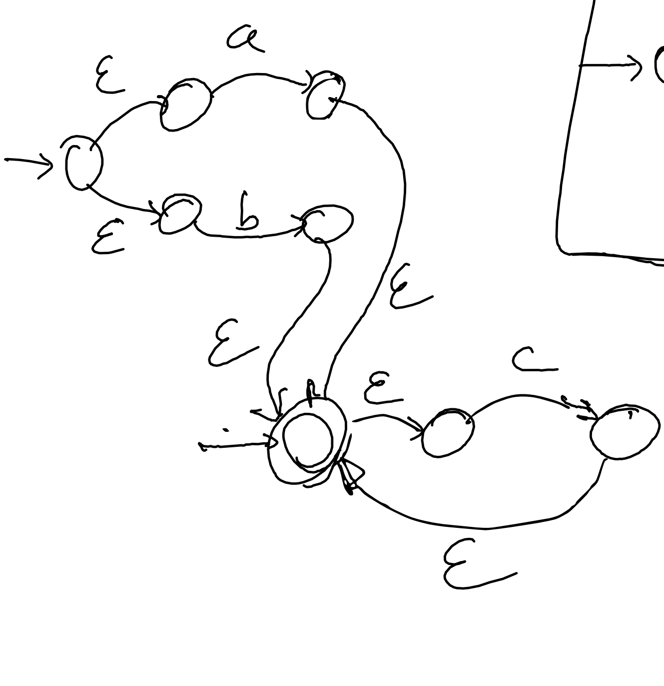


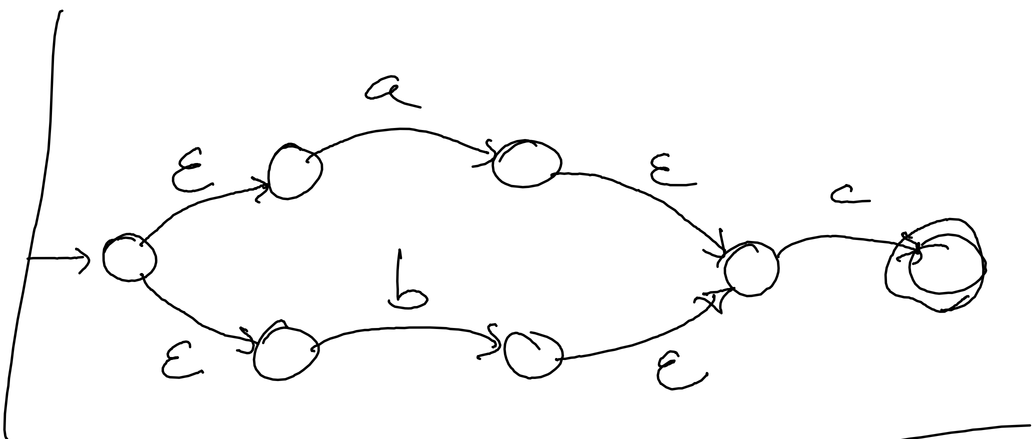
opération rationnelle.

expressions rationnelles

automates.

$(a|b).c^*$





$(a|b).c$

Elimination des ε-transitions

$(a|b).c^*$