X5I0050 - Langages et automates
Équivalence entre formalismes

D. Béchet & T. Sadiki

Université de Nantes & Université Internationale de Rabat

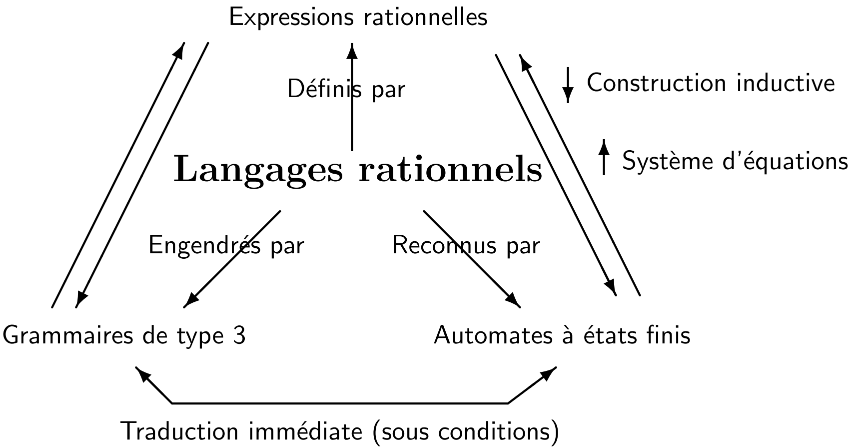
10 septembre 2014

Les langages rationnels

Formalismes décrivant les langages rationnels :

- Expressions rationnelles (ou définitions ensemblistes)
- Grammaires rationnelles (grammaires de type 3)
- Automates finis (déterministes ou non, etc)

Équivalence des formalismes pour les langages rationnels



Expression rationnelle $\equiv$ grammaire de type 3

Rappel : $G = (VT, VN, S, R)$ est rationnelle (de type 3)

$$\text{ssi } \forall \alpha \to \beta \in R : \begin{cases} \alpha \in VN \\ \beta \in VT^{+} \cup VT^{*} \cdot VN \end{cases}$$

Si le langage contient le mot vide :

$$\begin{cases} S \to \epsilon \in R \\ \forall \alpha \to \beta \in R,\ S \text{ n'apparaît pas dans } \beta \end{cases}$$

Théorème : Langages engendrés par une grammaire rationnelle $L(G_{rat}) =$ Langages rationnels (L_{rat})

Preuve :

1. Montrer $L_{rat} \subseteq L(G_{rat})$ en construisant une grammaire générant un langage vérifiant une définition inductive donnée
2. Montrer $L(G_{rat}) \subseteq L_{rat}$ en définissant le langage engendré par une grammaire rationnelle donnée

$L_{rat} \subseteq L(G_{rat})$

- $\emptyset$ engendré par $(V, \{S\}, S, \emptyset)$
- $\{\epsilon\}$ engendré par $(V, \{S\}, S, \{S \rightarrow \epsilon\})$
- $\{m\}$ engendré par $(V, \{S\}, S, \{S \rightarrow m\})$
- Induction :

$$\text{Soient} \left\{ \begin{array}{l} L1: \;\; G_1 = (V, VN_1, S_1, R_1) \\ L2: \;\; G_2 = (V, VN_2, S_2, R_2) \quad \text{avec } VN_1 \cap VN_2 = \emptyset \\ S \notin VN_1 \cup VN_2 \end{array} \right.$$

  - $L_1 \cup L_2 : \;\; G = (V, VN_1 \cup VN_2 \cup \{S\}, S, R_1 \cup R_2 \cup \{S \rightarrow S_1 \mid S_2\})$
  - $L_1 \cdot L_2 : \;\; G = (V, VN_1 \cup VN_2, S_1,$
    $R_2 \cup \{\alpha \rightarrow mX \text{ tq } \alpha \rightarrow mX \in R_1, X \in VN_1\}$
    $\cup \{\alpha \rightarrow mS_2 \text{ tq } \alpha \rightarrow m \in R_1, m \in V^*\})$
  - $L_1^* : \;\; G = (V, VN_1 \cup \{S, S_2\}, S,$
    $R_1 \cup \{S \rightarrow S_2 \mid \epsilon, \; S_2 \rightarrow S_1\}$
    $\cup \{\alpha \rightarrow mS_2 \text{ tq } \alpha \rightarrow m \in R_1, m \in V^*\})$

Attention : le type de G peut ne pas être 3 si $S_1 \rightarrow \epsilon \in R_1$ ou $S_2 \rightarrow \epsilon \in R_2$

$\Rightarrow$ Il faut aussi éliminer les règles $X \rightarrow \epsilon$ lorsque $X \neq S$

$L(G_{rat}) \subseteq L_{rat}$

Soit $G = (VT, VN, S, R)$ une grammaire de type 3

- Pour chaque $z \in VN$, on définit L_Z le langage des mots engendrés à partir de Z (à la place de l'axiome S) :
  - Soit $Z \to m_1 \mid m_2 \mid \cdots \mid m_k \mid w_1 Z_1 \mid w_2 Z_2 \mid \cdots \mid w_l Z_l \in R$ l'ensemble des règles ayant pour partie gauche le symbole non-terminal Z, avec $m_i \in VT^+$, $w_i \in VT^*$ et $Z_i \in VN$
  - Alors, $L_Z = \{m_1\} \cup \{m_2\} \cup \cdots \cup \{m_k\}$
    $\cup(\{w_1\} \cdot L_{Z_1}) \cup (\{w_2\}.L_{Z_2}) \cup \cdots \cup (\{w_l\}.L_{Z_l})$

$\Rightarrow$ Système d'équations ensemblistes sur les L_{Z_i}

$L(G_{rat}) \subseteq L_{rat}$

Théorème (voir le lemme d'Arden) : tout système d'équations de la forme

$$L_1 = A_1 \cup B_{11}L_1 \cup \cdots \cup B_{1n}L_n$$
$$\cdots$$
$$L_n = A_1 \cup B_{n1}L_1 \cup \cdots \cup B_{nn}L_n$$

admet une solution unique pour les langages $L_1, \ldots, L_n$ qui sont tous des langages rationnelles (sous la condition que $\epsilon \notin B_{ij}$)

Attention : Il faut que $\epsilon \notin B_{ij}$ donc que $w_i \neq \epsilon$ c'est-à-dire qu'il n'existe pas de règle $Z \rightarrow Z_i \in R$

$\Rightarrow$ Il faut éliminer au préalable les règles $\alpha \rightarrow X, X \in VN$

Automate fini $\equiv$ grammaire de type 3

Soit un automate fini $AF = (V, Q, I, F, T)$ avec contraintes :

- V : **alphabet** des symboles d'entrée
- $Q = \{q_1, \ldots, q_n\}$: **les états**
- $I = \{q_{i_0}\} \subseteq Q$: **l'état initial** (supposé unique)
- $F = \{q_f\} \subseteq Q$: **l'état final** (supposé unique)
- T : **transitions** (q_i, x, q_j), $q_i \in Q - \{q_f\}, q_j \in Q$, $x \in V$
- $q_{i_0} \neq q_f$ (le langage ne contient pas le mot vide)

Soit une grammaire de type 3 $G = (VT, VN, S, R)$ avec contraintes :

- pas de règle $S \rightarrow \epsilon$ (le langage ne contient pas le mot vide)
- Règles de la forme $X \rightarrow a$ ou $X \rightarrow aY$ avec $X, Y \in VN$ et $a \in VT$ (le membre droit des règles ne peut par être de la forme $X \rightarrow a_1 \cdots a_n$ ou $X \rightarrow a_1 \cdots a_n Y$ avec $n \geq 2$ ou de la forme $X \rightarrow Y$)

Remarque : les restrictions sur la forme des automates et des grammaires ne sont pas difficiles à obtenir.

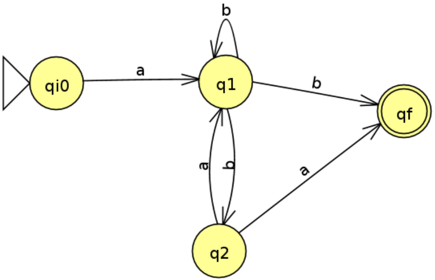
Automate fini $\equiv$ grammaire de type 3

On a correspondance entre AF et G si :

- $V = VT$
- $Q = VN \cup \{q_f\}$ ou $VN = Q - \{q_f\}$
- $q_{i_0} = S$
- $R = \{q_i \rightarrow xq_j \text{ tq } (q_i, x, q_j) \in Q \text{ et } q_j \notin F\}$
  $\cup \{q_i \rightarrow x \text{ tq } (q_i, x, q_j) \in Q \text{ et } q_j \in F\}$
- $T = \{(q_i, x, q_j) \text{ tq } q_i \rightarrow xq_j \in R,\ q_i, q_j \in VN,\ x \in VT\}$
  $\cup \{(q_i, x, q_f) \text{ tq } q_i \rightarrow x \in R,\ q_i \in VN,\ x \in VT\}$

Automate fini $\equiv$ grammaire de type 3

Exemple

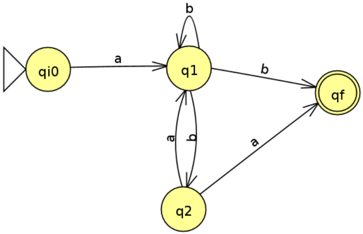


Grammaire $G = (\{a, b\}, \{S, Q_1, Q_2\}, S, R)$ avec :

$$R = \left\{ \begin{array}{lcl} S & \rightarrow & aQ_1 \\ Q_1 & \rightarrow & bQ_1 \mid bQ_2 \mid b \\ Q_2 & \rightarrow & aQ_1 \mid a \end{array} \right\}$$

Automate fini / grammaire de type 3 $\Rightarrow$ expression rationnelle

Exemple



ou $\left\{ \begin{array}{lcl} S & \rightarrow & aQ_1 \\ Q_1 & \rightarrow & bQ_1 \mid bQ_2 \mid b \\ Q_2 & \rightarrow & aQ_1 \mid a \end{array} \right\}$

Système d'équations (linéaires droites) : $R = \left\{ \begin{array}{lcl} Q_{i_0} & = & aQ_1 \\ Q_1 & = & bQ_1 + bQ_2 + b \\ Q_2 & = & aQ_1 + a \end{array} \right.$

Une solution unique : $R = \left\{ \begin{array}{lcl} Q_{i_0} & = & a(ba^?)^+ \\ Q_1 & = & (ba^?)^+ \\ Q_2 & = & a(ba^?)^* \end{array} \right.$