Fiche récap automates et expressions régulières

Contents

I) Expressions régulières / rationnelles

On se donne un alphabet Σ.

Définition 1 [Expression régulière]

> On défini inductivement une expression régulière comme :
> - $\emptyset$ l'expression vide
> - ε le mot vide
> - $x, x \in \Sigma$ une lettre
>
> Pour e_1, e_2 deux expressions régulières, on ajoute :
> - $e_1 + e_2$ ou $e_1 \mid e_2$ l'union
> - $e_1 e_2$ ou $e_1 \cdot e_2$ la concaténation
> - $e_1^\star$ l'étoile

Définition 2 [Language associé à une expression]

> On défini le language associé à une expression régulière inductivement
> - $L(\emptyset) = \emptyset$
> - $L(\varepsilon) = \{\varepsilon\}$
> - $\forall x \in \Sigma, L(x) = \{x\}$
>
> Pour e_1, e_2 deux expressions régulières,
> - $L(e_1 \mid e_2) = L(e_1) \cup L(e_2)$
> - $L(e_1 e_2) = \{m_1 m_2 \mid m_1 \in L(e_1), m_2 \in L(e_2)\}$
> - $L(e_1^\star) = \{m_1 m_2 ... m_n \mid m_1, ..., m_n \in L(e_1)\}$ (à noter que $\varepsilon \in L(e_1^\star)$)

On appelle language *régulier* un language reconnu par une expression régulière.

II) Automates

II.1) Déterministes

Définition 3 [Automate déterministe]

> Un automate A est un quintuplet $(Q, \Sigma, q_0, F, \delta)$, où :
> - Q est un ensemble fini d'*états*
> - Σ est un alphabet
> - $q_0 \in Q$ est l'*état initial* de l'automate
> - $F \subset Q$ est l'ensemble des *états acceptants* de l'automate
> - $\delta : Q \times \Sigma \to Q \cup \{\emptyset\}$ est l'ensemble des transitions

Définition 4 [$\delta^\star$]

On défini inductivement la fonction $\delta^\star : Q \times \Sigma^\star \to Q$ de la manière suivante :
- $\forall q \in Q, \delta^\star(q, \varepsilon) = q$
- $\forall q \in Q, xu \in \Sigma^+, \delta^\star(q, xu) = \delta^\star(\delta(q, x), u)$

Définition 5 [Language reconnu par un automate]

On défini le language reconnu par un automate $A = (Q, \Sigma, q_0, F, \delta)$ comme :

$$L(A) = \{u \in \Sigma^\star \mid \delta^\star(q_0, u) \in F\}$$

II.2) Automates non déterministes

Définition 6 [Automate non déterministe]

Un automate A est un quintuplet $(Q, \Sigma, Q_0, F, \delta)$, où :
- Q est un ensemble fini d'*états*
- Σ est un alphabet
- $Q_0 \subset Q$ est l'ensemble des *états initiaux* de l'automate
- $F \subset Q$ est l'ensemble des *états acceptants* de l'automate
- $\delta : Q \times \Sigma \to \mathscr{P}(Q)$ est l'ensemble des transitions

Définition 7 [$\delta^\star$]

On défini inductivement la fonction $\delta^\star : \mathscr{P}(Q) \times \Sigma^\star \to \mathscr{P}(Q)$ de la manière suivante :
- $\forall E \subset Q, \delta^\star(E, \varepsilon) = E$
- $\forall E \subset Q, xu \in \Sigma^+, \delta^\star(E, xu) = \delta^\star(\delta(E, x), u)$

Définition 8 [Language reconnu par un automate]

On défini le language reconnu par un automate $A = (Q, \Sigma, q_0, F, \delta)$ comme :

$$L(A) = \{u \in \Sigma^\star \mid \delta^\star(Q_0, u) \cap F \neq \emptyset\}$$

II.3) Propriétés

Propriété 9 [Determination]

Soit $\mathscr{L}_{\text{dfa}}$ l'ensemble des languages reconnaissables par un automate déterministe.
Soit $\mathscr{L}_{\text{nfa}}$ l'ensemble des languages reconnaissables par un automate non-déterministe.

Alors : $\mathscr{L}_{\text{dfa}} = \mathscr{L}_{\text{nfa}}$

Définition 10 [Language rationnel]

On appelle language rationnel tout language L tel que $\exists A$ un automate, $L = L(A)$

Définition 11 [Automate des parties]

Soit $A_n = (Q_n, \Sigma, Q_0^n, F_n, \delta_n)$ un automate non deterministe.

On défini A_p l'automate des parties correspondant à A_n, de la manière suivante :

- $Q_p = \mathscr{P}(Q_n)$ (les parties de Q_n)
- $\Sigma = \Sigma$
- $q_0^p = Q_0^n$
- $F_p = \{E \mid E \subset Q_n \text{ et } E \cap F_n \neq \emptyset\}$
- $\delta_{p(E,x)} = \bigcup_{q \in E} \delta_{n(q,x)}$

Propriété 12 []

> A_n et A_p reconnaissent le même language.

Opération sur les languages rationnels 13 []

> Soit L_1, L_2 deux languages rationnels. Alors $L_1 \cup L_2, L_1 L_2, L_1 \cap L_2, L_2 \setminus L_2$ et $L_1^\star$ le sont aussi.

On le montre en construisant les automates correspondants, par exemple à l'aide de l'automate des parties.

III) Lien entre les deux

Théorème 14 [De Kleene]

> Les languages rationnels sont exactement les mêmes que les languages réguliers.

III.1) Passage automate -> expression régulière

On étiquette les transitions par des expressions régulières plutôt que des lettres, et on supprime les sommets un à un.

III.2) Passage expression régulière -> automate

Exemple : $(a + ba^\star b)^\star ba^\star$

On construit l'automate de Glushkov :

1. On *linéarise* l'expression régulière, c'est à dire qu'on numérote chaque lettre. On note $\Sigma' = \{a_1, a_2, a_3, b_1, b_2, b_3\}$ le nouvel alphabet.

$$(a_1 + b_1 a_2^\star b_2)^\star b_3 a_3^\star$$

2. On trouve l'ensemble des préfixes et suffixes de 1 lettre, et des facteurs de 2 lettres, possibles.

$$\text{premiers} : a_1, b_1, b_3$$

$$\text{derniers} : a_3, b_3$$

$$\text{facteurs} : a_1 a_1, a_1 b_1, a_1 b_3, b_1 a_2, b_1 b_2, a_2 a_2, a_2 b_2, b_2 a_1, b_2 b_1, b_2 b_3, b_3 a_3, a_3 a_3$$

3. On construit l'automate correspondant au language linéarisé $A = (Q = \Sigma \cup \{\varepsilon\}, \Sigma, \varepsilon, \text{derniers}, \delta)$, où :
   - $\forall x \in \text{premiers}, \delta(\varepsilon, x) = x$
   - $\forall xy \in \text{facteurs}, \delta(x, y) = y$
   - δ n'est pas définie sinon.

4. On supprime la numérotation dans les transitions du graphe pour avoir l'automate final.

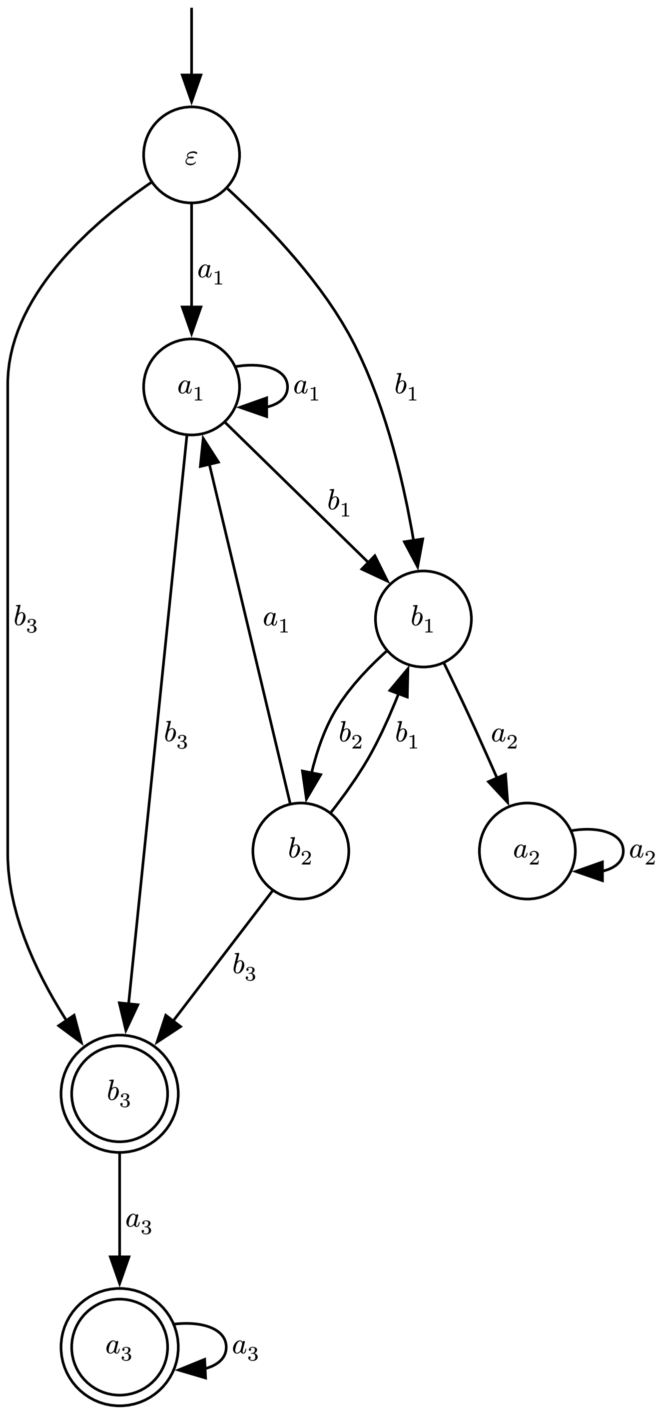

Figure 1: Automate de l’étape 3

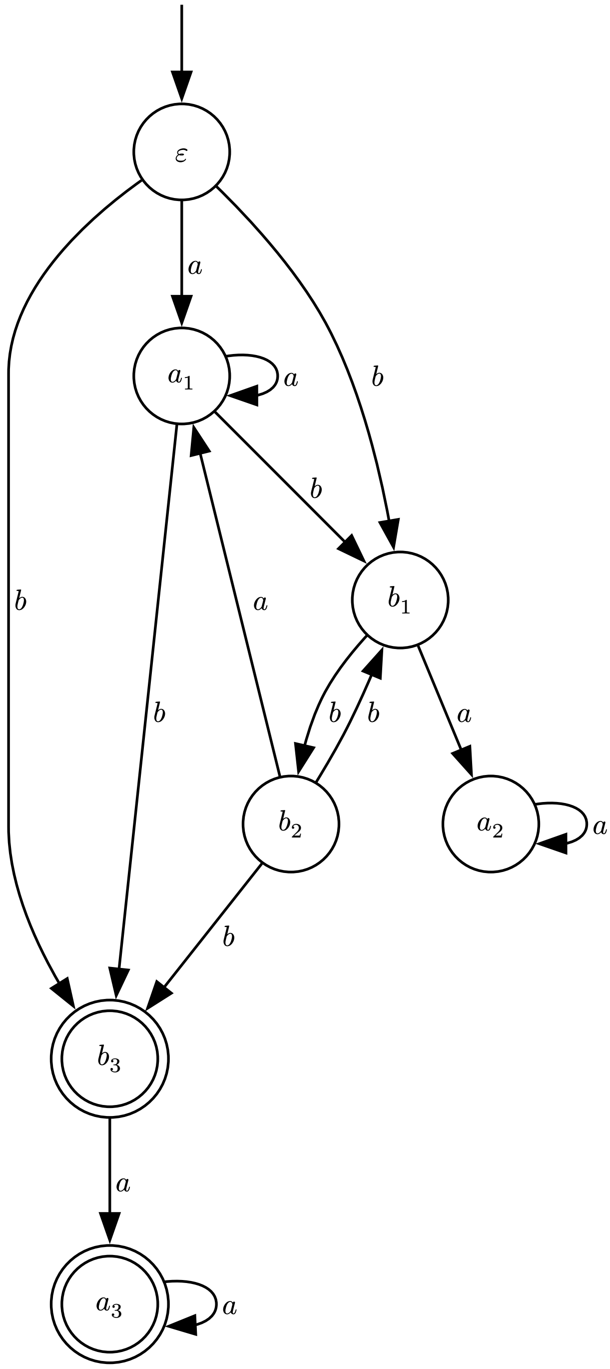

Figure 2: Automate de l’étape 4

À noter qu’on obtient un automate non-déterministe, qui contient autant d’état que la taille de l’expression régulière source.