Langages rationnels

FOND1 - ENS Rennes

Définition 1 (alphabet)**.** *Un* alphabet *est un ensemble de lettres. On suppose que les alphabets sont finis.*

Définition 2 (mot fini)**.** *Un* mot fini *sur un alphabet* Σ *est une suite finie de lettres de* Σ*. Le mot vide se note* ϵ*. On note* Σ^* *l'ensemble de tous les mots finis sur* Σ*.*

Définition 3 (concaténation)**.** *La concaténation de* u *et* v *se note* $u.v$*.*

Définition 4 (langage)**.** *Un* langage *sur un alphabet* Σ *est un sous-ensemble de* Σ^**.*

1 Opérations rationnelles

Définition 5 (Union)**.** $L_1 \cup L_2 = \{w \in \Sigma^* \mid w \in L_1 \text{ ou } w \in L_2\}$.

Définition 6 (Concaténation)**.** $L_1 \bullet L_2 = \{u \bullet v \mid u \in L_1, v \in L_2\}$.

Définition 7 (Puissance)**.** L^n *est défini par récurrence sur* n *:*
- $L^0 = \{\epsilon\}$;
- $L^{n+1} = L \bullet L^n$.

Définition 8 (Étoile de Kleene)**.** $L^* = \bigcup_{n \in \mathbb{N}} L^n$.

Proposition 9. L^* *est le plus petit langage (pour l'inclusion) contenant* ϵ*, incluant* L *et stable par concaténation.*

2 Expressions rationnelles

2.1 Syntaxe

Définition 10 (expression rationnelle)**.** *Soit* Σ *un alphabet. L'ensemble des expressions rationnelles* $ER(\Sigma)$ *est un langage sur l'alphabet* $\Sigma \cup \left\{ \bullet, \cup, *, \emptyset, \epsilon, (,) \right\}$ *est le plus petit langage (pour l'inclusion) tel que :*
- $\emptyset \in ER(\Sigma)$, $\quad \epsilon \in ER(\Sigma)$, $\quad \Sigma \subseteq ER(\Sigma)$;
- *Si* $\alpha_1, \alpha_2 \in ER(\Sigma)$*, alors*
  - $(\alpha_1 \cup \alpha_2) \in ER(\Sigma)$;
  - $(\alpha_1 \bullet \alpha_2) \in ER(\Sigma)$;
  - $\alpha_1^* \in ER(\Sigma)$.

Remarque 11. $ER(\Sigma)$ *est engendré par la grammaire suivante :*

$$\mathtt{E} ::= \emptyset \mid \epsilon \mid \mathtt{a} \mid (\mathtt{E} \bullet \mathtt{E}) \mid (\mathtt{E} \cup \mathtt{E}) \mid \mathtt{E}^*$$

où a *parcourt* Σ*.*

2.2 Sémantique

Définition 12 (sémantique d'une expression rationnelle)**.** *On définit* $L : ER(\Sigma) \to 2^{\Sigma^*}$ *par induction :*
- $L(\emptyset) = \emptyset$;
- $L(\epsilon) = \{\epsilon\}$;
- $L(\mathtt{a}) = \{a\}$ *pour tout* $a \in \Sigma$;
- $L(\alpha_1 \cup \alpha_2) = L(\alpha_1) \cup L(\alpha_2)$;
- $L(\alpha_1 \bullet \alpha_2) = L(\alpha_1) \bullet L(\alpha_2)$;
- $L(\alpha^*) = L(\alpha)^*$.

Définition 13 (langage rationnel)**.** *On appelle* langage rationnel *un langage qui est dans* $L(ER(\Sigma))$*. L'ensemble des langages rationnels est noté* $Rat(\Sigma)$*.*

3 Automates finis

Définition 14 (automate fini)**.** *Un* automate fini (non déterministe avec ϵ-transitions) *est un quintuplet* $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ *où*

- Q *est un ensemble fini non vide d'états ;*
- Σ *est un alphabet (fini) ;*
- $\delta \subseteq Q \times \Sigma_\epsilon \times Q$ *est une relation de transition ;* *où* $\Sigma_\epsilon = \Sigma \cup \{\epsilon\}$.
- $q_0 \in Q$ *est l'état initial ;*
- $F \subseteq Q$ *est l'ensemble des états finaux.*

Définition 15 (exécution)**.** *Soit* w *un mot. Une* exécution *de* $\mathcal{A}$ *sur le mot* w *de l'état* q *vers l'état* q' *est une séquence* $(r_0, y_1, \ldots, y_n, r_n)$ *avec :*

- $r_0 = q$ *et* $r_n = q'$ *;*
- $w = y_1 y_2 \ldots y_n$ *avec, pour tout* $i \in \{1, \ldots, n\}$, $y_i \in \Sigma_\epsilon$ *;*
- *Pour tout* $i \in \{0, \ldots, n-1\}$, $(r_i, y_i, r_{i+1}) \in \delta$.

L'exécution est notée $r_0 \xrightarrow{y_1} r_1 \cdots r_{n-1} \xrightarrow{y_n} r_n$.

On note aussi $q \xrightarrow{w} q'$ pour dire qu'il existe une exécution de $\mathcal{A}$ sur le mot w de q vers q'.

Définition 16 (exécution acceptante)**.** *Quand* $q_0 \xrightarrow{w} q_f$ *avec* $q_f \in F$*, on dit que l'exécution est* acceptante*.*

Définition 17 (langage accepté)**.** *Le* langage accepté par $\mathcal{A}$*, noté* $L(\mathcal{A})$*, est défini par*

$$L(\mathcal{A}) = \left\{ w \in \Sigma^* \mid \exists q_f \in F, q_0 \xrightarrow{w} q_f \right\}.$$

Définition 18 (langage reconnaissable)**.** *Un langage* L *est* reconnaissable *s'il existe un automate* $\mathcal{A}$ *tel que* $L = L(\mathcal{A})$*. La classe des langages reconnaissable sur l'alphabet* Σ *est noté* $Rec(\Sigma)$*.*

4 Théorème de Kleene

Théorème 19 (Kleene, 1956)**.** $Rat(\Sigma) = Rec(\Sigma)$*.*

4.1 Construction de Thomson : rat $\Rightarrow$ rec

Proposition 20. *Soit* e *une expression rationnelle. Il existe un automate fini* $\mathcal{A}_e$ *non-déterministe avec* ϵ*-transitions tel que* $L(e) = L(\mathcal{A}_e)$*. De plus, le nombre d'états de* $\mathcal{A}_e$ *borné par* $2|e|$*.*

4.2 Brzozowski-McCluskey : rec $\Rightarrow$ rat

Définition 21 (automate fini non déterministe généralisé)**.** *Un* automate fini non déterministe généralisé *est un quintuplet* $\mathcal{A} = (Q, \Sigma, \delta, q_0, q_f)$ *où*

- Q *est un ensemble fini non vide d'états ;*
- Σ *est un alphabet (fini) ;*
- $\delta : (Q \setminus \{q_f\}) \times (Q \setminus \{q_0\}) \to ER(\Sigma)$ *est la fonction de transition ;*
- $q_0 \in Q$ *est l'état initial ;*
- $q_f \in Q$, $q_f \neq q_0$ *est l'état final.*

Définition 22 (exécution d'un automate non déterministe généralisé)**.** *Soit* w *un mot. Une* exécution *de* $\mathcal{A}$ *sur le mot* w *de l'état* q *vers l'état* q' *est une séquence* $(r_0, y_1, \ldots, y_n, r_n)$ *avec :*

- $r_0 = q$ *et* $r_n = q'$ *;*
- $w = y_1 y_2 \ldots y_n$ *avec, pour tout* $i \in \{1, \ldots, n\}$, $y_i \in \Sigma^*$ *;*
- *Pour tout* $i \in \{0, \ldots, n-1\}$, $y_i \in L(\delta(r_i, r_{i+1}))$.

De la même façon, l'exécution est notée $r_0 \xrightarrow{y_1} r_1 \cdots r_{n-1} \xrightarrow{y_n} r_n$. On note $L(\mathcal{A})$ le langage accepté par $\mathcal{A}$: c'est l'ensemble des mots w tel qu'il existe une exécution de $\mathcal{A}$ sur le mot w de l'état q_0 à q_f.

Proposition 23. *Soit* $\mathcal{A}$ *un automate non-déterministe avec* ϵ*-transitions. Il existe une expression rationnelle* $e_{\mathcal{A}}$ *telle que* $L(e_{\mathcal{A}}) = L(\mathcal{A})$*.*

5 Lemme de pompage

Lemme 24. *Soit* L *un langage rationnel. Il existe un entier* N *tel que pour tout mot* $u \in L$*, si* $|u| \geq N$*, alors* u *s'écrit* $u = vtw$ *avec* $|t| > 0$ *et pour tout* $k \in \mathbb{N}$, $vt^k w \in L$*.*