TD 1 - Langages réguliers

définitions

Longueur (nombre de symboles) d'un mot :

$$|\varepsilon| = 0 \qquad\qquad |aw| = 1 + |w|$$

Nombre d'occurences d'un symbole :

$$|\varepsilon|_a = 0 \qquad\qquad |bw|_a = \begin{cases} 1 + |w|_a & si\ a = b \\ |w|_a & si\ a \neq b. \end{cases}$$

Image miroir (inverse l'ordre des lettres) :

$$\overline{\varepsilon} = \varepsilon \qquad\qquad \overline{aw} = \overline{w}a$$

Concaténation et étoile de langages :

$$L_1 \cdot L_2 := \{u_1 \cdot u_2 \mid u_1 \in L_1,\ u_2 \in L_2\}$$
$$L^0 := \{\varepsilon\} \qquad\qquad L^{n+1} := L \cdot L^n$$
$$L^\star := \bigcup_{n \in \mathbb{N}} L^n = \{u_1 \cdots u_n \mid n \in \mathbb{N},\ \forall 1 \leqslant i \leqslant n,\ u_i \in L\}.$$

Exercice 1. Langages réguliers

Montrer que les langages suivants sont réguliers.

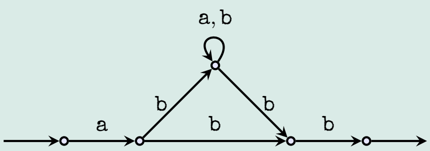
Solution :

Pour chacune des questions suivantes, grâce au théorème de Kleene, on peut au choix donner un automate ou une expression régulière reconnaissant le langage demandé pour montrer qu'il est régulier. Dans cette correction, on donne dans chaque cas les deux (ce qui est superflu, un seul des deux formalismes suffit).

1. $\{w \in \{\mathtt{a},\mathtt{b}\}^\star \mid w$ commence par $\mathtt{ab}$ et finit par $\mathtt{bb}\}$;

   Solution :

   Voici un automate reconnaissant le langage en question :

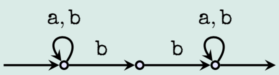
   

   Et voici une expression régulière : $(\mathtt{a}\cdot\mathtt{b}) \cup (\mathtt{a}\cdot\mathtt{b}\cdot(\mathtt{a}\cup\mathtt{b})^\star\cdot\mathtt{b}\cdot\mathtt{b})$.

2. $\{w \in \{\mathtt{a},\mathtt{b}\}^\star \mid w$ a (au moins) une occurence de $\mathtt{bb}\}$;

   Solution :

   Voici un automate reconnaissant le langage en question :

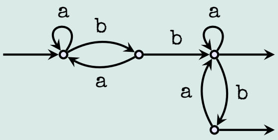
   

   Et voici une expression régulière : $(\mathtt{a}\cup\mathtt{b})^\star\cdot\mathtt{b}\cdot\mathtt{b}\cdot(\mathtt{a}\cup\mathtt{b})^\star$.

3. $\{w \in \{\mathtt{a},\mathtt{b}\}^\star \mid w$ a exactement une occurence de $\mathtt{bb}\}$;

Solution :

Voici un automate reconnaissant le langage en question :

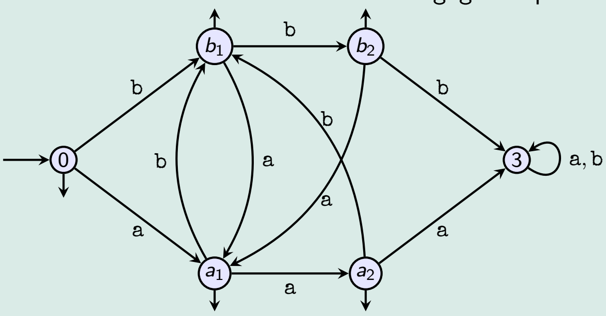


Et voici une expression régulière : $(\mathtt{a}\cup\mathtt{ba})^\star\cdot\mathtt{bb}\cdot(\mathtt{a}\cup\mathtt{ba})^\star\cdot(\varepsilon\cup\mathtt{b})$.

4. $\{w \in \{\mathtt{a},\mathtt{b}\}^\star \mid w$ ne contient pas 3 occurences successives de la même lettre$\}$;

Solution :

Voici un automate reconnaissant le langage en question :

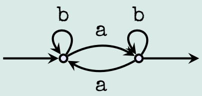


Et voici une expression régulière : $(\varepsilon\cup\mathtt{b}\cup\mathtt{bb})\cdot(\mathtt{a}\cdot(\varepsilon\cup\mathtt{a})\cdot(\mathtt{b}\cup\mathtt{bb}))^\star\cdot(\varepsilon\cup\mathtt{a})$.

5. $\{w \in \{\mathtt{a},\mathtt{b}\}^\star \mid w$ contient un nombre impair de $\mathtt{a}\}$;

Solution :

Voici un automate reconnaissant le langage en question :

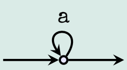


Et voici une expression régulière : $\mathtt{b}^\star\cdot\mathtt{a}\cdot(\mathtt{b}^\star\cdot\mathtt{a}\cdot\mathtt{b}^\star\cdot\mathtt{a})^\star\cdot\mathtt{b}^\star$.

6. $\{w \in \{\mathtt{a},\mathtt{b}\}^\star \mid w$ ne contient pas la lettre $\mathtt{b}\}$.

Solution :

Voici un automate reconnaissant le langage en question :



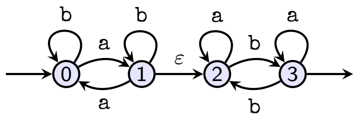
Et l'expression régulière idoine : $\mathtt{a}^\star$.

Exercice 2. Élimination des ε-transitions

Un automate avec ε-transitions (ou automate asynchrone) est une structure

$$\mathscr{A} := \langle Q, \Sigma, \Delta, I, F\rangle$$

où Q est un ensemble fini d'états, Σ est un alphabet fini, $I, F \subseteq Q$ sont les ensembles d'états respectivement initiaux et finaux, et $\Delta \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times Q$ est l'ensemble des transitions. Voici un exemple :



Une exécution dans un tel automate est une séquence

$$q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} q_n$$

telle que $\forall 1 \leqslant i \leqslant n$, $q_{i-1} \xrightarrow{a_i} q_i \in \Delta$, et $\langle q_0, q_n \rangle \in I \times F$. Le mot reconnu par l'exécution est obtenu à partir de la séquence $a_1 \cdots a_n$ en effaçant les εs, et en ne gardant que les symboles de l'alphabet Σ. Le langage de l'automate est l'ensemble des mots reconnus par des exécutions de l'automate.

Le langage de l'automate donné en exemple ci-dessus est l'ensemble des mots $w \in \{\mathtt{a}, \mathtt{b}\}$ qui peuvent se décomposer en $w = uv$, où u a un nombre impair de $\mathtt{a}$, et v a un nombre impair de $\mathtt{b}$. En effet, toute exécution de cette machine se décompose comme suit : $0 \xrightarrow{u}{}^{*} 1 \xrightarrow{\varepsilon} 2 \xrightarrow{v}{}^{*} 3$. L'exécution $0 \xrightarrow{u}{}^{*} 1$ nous assure que le mot u contient un nombre impair de $\mathtt{a}$ (voir la question 5 de l'exercice 1), et de manière similaire la fin de l'exécution vérifie que v contient un nombre impair de $\mathtt{b}$.

Question. Montrer que les langages reconnus par les automates asynchrones sont exactement les langages réguliers.

Solution :

Il faut montrer une équivalence, c'est à dire deux implications :

— Pour tout automate, il existe un automate asynchrone qui reconnait le même langage.

— Pour tout automate asynchone, il existe un automate qui reconnait le même langage.

La première implication est triviale : tout automate est en particulier un automate asynchrone (un tel automate n'est pas *forcé* d'utiliser des ε-transitions).
Pour la deuxième, soit $\mathscr{A}$ un automate asynchrone. On construit un nouvel automate sans ε-transitions comme suit :

$$\mathscr{A}' := \left\langle Q, \Sigma, \Delta', I, F' \right\rangle$$

avec :

$$\begin{aligned} \Delta' &:= \{p \xrightarrow{a} q \mid a \in \Sigma \text{ et } \exists r \in Q : \ p \xrightarrow{\varepsilon}{}^{*} r \text{ et } r \xrightarrow{a} q \in \Delta\} \\ F' &:= \{p \in Q \mid \exists q : \ p \xrightarrow{\varepsilon}{}^{*} q \in F\} \end{aligned}$$

Vérifions que cet automate reconnait le même langage que $\mathscr{A}$. Soit $w = a_1 \dots a_n$ un mot sur l'alphabet Σ. Alors $w \in \mathcal{L}(\mathscr{A})$ si et seulement si il existe une exécution

$$I \ni q_0 \xrightarrow{b_1} q_1 \cdots \xrightarrow{b_m} q_m \in F$$

telle que le mot obtenu à partir de la séquence $b_1 \cdots b_n$ en effaçant les ε est exactement w. Autrement dit, si on note k_i la longueur de la séquence d'ε entre a_{i-1} et a_i, la séquence des b_is se réécrit :

$$b_1 \cdots b_n = \varepsilon^{k_0} a_1 \varepsilon^{k_1} a_2 \cdots \varepsilon^{k_{n-1}} a_n \varepsilon^{k_n}.$$

De manière équivalente, on peut donc dire que $w \in \mathcal{L}(\mathscr{A})$ si et seulement si il existe une exécution de la forme suivante :

$$I \ni q_0 \xrightarrow{\varepsilon} q_0^1 \cdots \xrightarrow{\varepsilon} q_0^{k_0} \xrightarrow{a_1} q_1 \cdots q_{n-1}^{k_{n-1}} \xrightarrow{a_n} q_n \xrightarrow{\varepsilon} q_n^1 \cdots \xrightarrow{\varepsilon} q_n^{k_n} \in F$$

Cela est équivalent à dire que

1. $q_0 \in I$;
2. pour tout $1 \leqslant i \leqslant n$ on a $q_{i-1} \xrightarrow{\varepsilon}{}^{*} q_{i-1}^{k_{i-1}}$ et $q_{i-1}^{k_{i-1}} \xrightarrow{a_i} q_i \in \Delta$, c'est à dire $q_{i-1} \xrightarrow{a_i} q_i \in \Delta'$;
3. $q_n \xrightarrow{\varepsilon}{}^{*} q_n^{k_n} \in F$, c'est à dire $q_n \in F'$.

Ces dernières conditions sont bien équivalente à $w = a_1 \cdots a_n \in \mathcal{L}(\mathscr{A}')$.

Exercice 3. Propriétés de clôture des langages réguliers

Montrer que les langages réguliers sont clos par :

1. union
2. intersection
3. concaténation
4. étoile (répétition)
5. image miroir.

Solution :

Soient $\mathscr{A}_1$ et $\mathscr{A}_2$ deux automates sur le même alphabet Σ, en notant $\mathscr{A}_i := \langle Q_i, \Sigma, \Delta_i, I_i, F_i \rangle$. On suppose que $Q_1 \cap Q_2 = \emptyset$.

1. $\mathscr{A}_\cup := \langle Q_1 \cup Q_2, \Sigma, \Delta_1 \cup \Delta_2, I_1 \cup I_2, F_1 \cup F_2 \rangle$.
2. $\mathscr{A}_\cap := \langle Q_1 \times Q_2, \Sigma, \Delta_\cap, I_1 \times I_2, F_1 \times F_2 \rangle$, avec :
$$\langle p_1, p_2 \rangle \xrightarrow{\mathtt{a}} \langle q_1, q_2 \rangle \in \Delta_\cap \Leftrightarrow \forall i \in \{1,2\},\ p_i \xrightarrow{\mathtt{a}} q_i \in \Delta_i.$$
3. $\mathscr{A}_; := \left\langle Q_1 \cup Q_2, \Sigma, \Delta_1 \cup \Delta_2 \cup \Delta', I_1, F_2 \right\rangle$, avec :
$$\Delta' := \{p \xrightarrow{\varepsilon} q \mid p \in F_1,\ q \in I_2\} = F_1 \times \{\varepsilon\} \times I_2.$$
4. $\mathscr{A}_\star := \left\langle Q_1, \Sigma, \Delta_1 \cup \Delta'', I_1, I_1 \right\rangle$, avec :
$$\Delta'' := \{p \xrightarrow{\varepsilon} q \mid p \in F_1,\ q \in I_1\} = F_1 \times \{\varepsilon\} \times I_1.$$
5. $\mathscr{A}_{\mathsf{miroir}} := \left\langle Q_1, \Sigma, \Delta_1', F_1, I_1 \right\rangle$, avec :
$$\Delta_1' := \{p \xrightarrow{\mathtt{a}} q \mid q \xrightarrow{\mathtt{a}} p \in \Delta_1\}.$$

Exercice 4. Lemme de l'étoile

Montrer que les langages suivants ne sont pas réguliers :

1. $L_1 := \{\mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N}\}$.

   Solution :

   Supposons L_1 régulier, et soit $w_n = \mathtt{a}^n\mathtt{b}^n$. D'après le lemme de l'étoile, il existe un entier $N \in \mathbb{N}$ tel que pour tout mot $w \in L_1$, si $|w| \geqslant N$, alors il existe des mots u_1, u_2, u_3 tels que :

   — $w = u_1 u_2 u_3$, $u_2 \neq \varepsilon$, et $|u_1 u_2| \leqslant N$;

   — $u_1 (u_2)^\star u_3 \subseteq L_1$, c'est à dire que pour tout entier $k \in \mathbb{N}$ on a $u_1 (u_2)^k u_3 \in L_1$.

   En particulier, comme $w_N \in L_1$, il doit exister une décomposition $w_N = u_1 u_2 u_3$ comme ci-dessus. Étant donné que $|u_1 u_2| \leqslant N$, nécessairement il existe des entiers k, ℓ tels que $k + \ell \leqslant N$, $l \neq 0$ et :
$$u_1 = \mathtt{a}^k \qquad u_2 = \mathtt{a}^\ell \qquad u_3 = \mathtt{a}^{N-k-\ell}\mathtt{b}^N$$
   D'après le lemme, on a $u_1 u_3 \in L_1$. Pourtant $u_1 u_3 = \mathtt{a}^{N-\ell}\mathtt{b}^N$, et comme $\ell \neq 0$ ce mot n'appartient pas à L_1 : on a une contradiction.

2. $L_2 := \left\{ w \in \{\mathtt{a}, \mathtt{b}\}^\star \;\middle|\; |w|_\mathtt{a} = |w|_\mathtt{b} \right\}$.

   Solution :

   Le mot w_N utilisé dans la question précédente fonctionne à nouveau ici : on a bien $w_N \in L_2$ car $|w|_\mathtt{a} = N = |w|_\mathtt{b}$, et $u_1 u_3 = \mathtt{a}^{N-\ell}\mathtt{b}^N \notin L_2$ car $|u_1 u_3|_\mathtt{a} = N - \ell \neq N = |u_1 u_3|_\mathtt{b}$.

3. $L_3 := \left\{ w \in \{\mathtt{a},\mathtt{b}\}^\star \mid w = \overline{w} \right\}$.

Solution :

Ici on choisit $w_N = \mathtt{a}^N\mathtt{b}\mathtt{a}^N$. On a des entiers k,ℓ tels que $k+\ell \leqslant N$, $\ell \neq 0$ et :

$$u_1 = \mathtt{a}^k \qquad u_2 = \mathtt{a}^\ell \qquad u_3 = \mathtt{a}^{N-k-\ell}\mathtt{b}\mathtt{a}^N$$

On obtient une contradiction avec $u_1u_3 = \mathtt{a}^{N-\ell}\mathtt{b}\mathtt{a}^N \notin L_3$ car $\overline{u_1u_3} = \mathtt{a}^N\mathtt{b}\mathtt{a}^{N-\ell} \neq u_1u_3$.

4. $L_4 := \left\{ ww \mid w \in \{\mathtt{a},\mathtt{b}\}^\star \right\}$.

Solution :

Ici on choisit $w_N = \mathtt{a}^N\mathtt{b}\mathtt{a}^N\mathtt{b}$. On a des entiers k,ℓ tels que $k+\ell \leqslant N$, $\ell \neq 0$ et :

$$u_1 = \mathtt{a}^k \qquad u_2 = \mathtt{a}^\ell \qquad u_3 = \mathtt{a}^{N-k-\ell}\mathtt{b}\mathtt{a}^N\mathtt{b}$$

On obtient une contradiction avec $u_1u_3 = \mathtt{a}^{N-\ell}\mathtt{b}\mathtt{a}^N\mathtt{b} \notin L_3$.