Synthèse sur les langages et automates

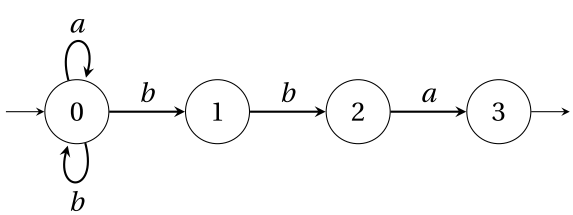
OPTION INFORMATIQUE - TP nº 4.4 - Olivier Reynet

À la fin de ce chapitre, je sais :

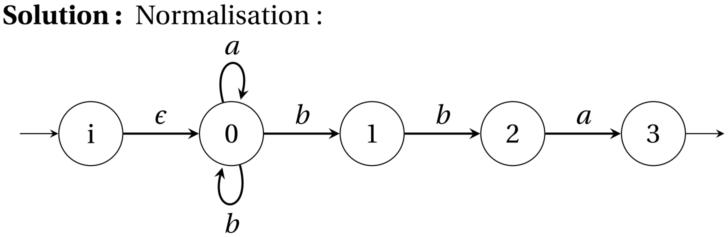
- ☞ Utiliser les définitions des mot et des langages
- ☞ Simplifier les expressions régulières
- ☞ Trouver le langage associé à une expression régulière
- ☞ Construire l'automate correspondant à un langage régulier
- ☞ Déterminiser un AFND
- ☞ Transformer une expression régulière en AFND avec Thompson ou Berry-Sethi
- ☞ Transformer un automate en une expression régulière
- ☞ Montrer qu'un langage n'est pas régulier

A Automate vers expression régulière

Pour les automates suivants, donner une expression régulière telle que le langage reconnaissable par l'automate est le langage dénoté par l'expression régulière. On éliminera les états après avoir normalisé l'automate en entrée et en sortie.

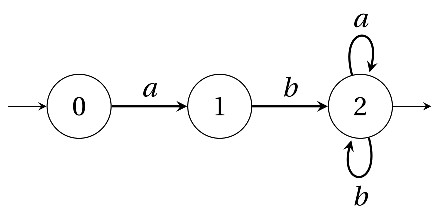
A1. 

Solution : Normalisation :

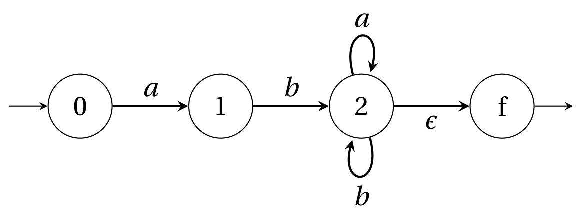


Par élimination d'états : $(a|b)^* bba$

A2.

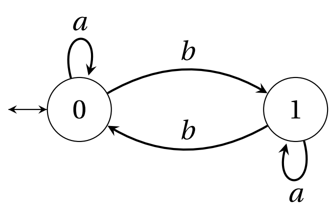


Solution : Normalisation :

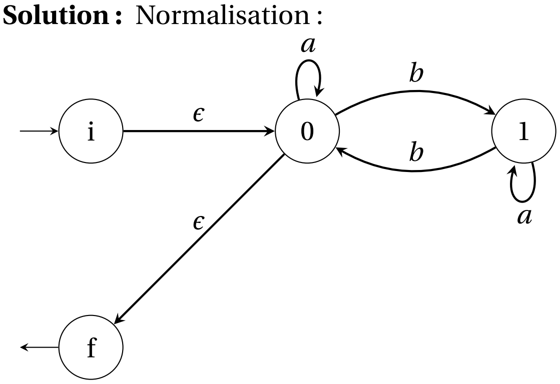


Par élimination d'états : $ab(a|b)^*$

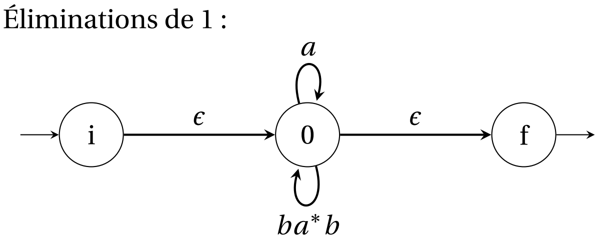
A3.



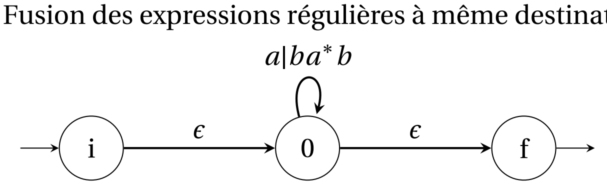
Solution : Normalisation :



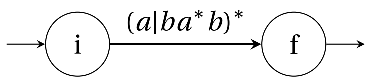
Éliminations de 1 :



Fusion des expressions régulières à même destination depuis 0 :

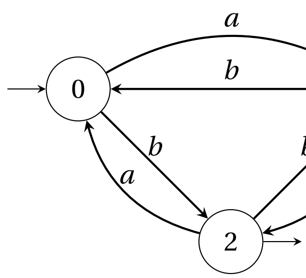


Boucle sur 0 :



Par élimination d'états : $(a|ba^*b)^*$

A4.



Solution : Par élimination d'états : $((a|bb)(b|aa))^*(b|((a|bb)a))$

B Expression régulière vers automate

Pour chaque expression régulière, appliquer :

1. l'algorithme de Berry-Sethi, trouver l'automate de Glushkov et le déterminer si besoin.
2. l'algorithme de Thompson (méthode compositionnelle), trouver l'automate et éliminer les transitions spontanées.

B1. $a|b$

B2. a^*b

B3. $(a|b)^*a$

B4. $(a|b)a^*b^*$

B5. $(a^*b)|(a(a|b)^*)$

B6. $(ba|a)^*ab$

B7. $(a|c)^*abb|(a|c)^*$

B8. $(ba|a)^*(a|b)c$

C Stabilité des langages réguliers

C1. Soit Σ un alphabet et $\mathcal{L}$ un langage régulier sur Σ. Montrer que le complémentaire de $\mathcal{L}$, $C(\mathcal{L}) = \overline{\mathcal{L}} = \{w, w \in \Sigma^* \text{ et } w \notin \mathcal{L}\}$, est régulier.

Solution : Soit $\mathcal{A}$ un automate fini qui reconnait le langage régulier $\mathcal{L}$. Un mot non reconnu par cet automate appartient à $\overline{\mathcal{L}}$. Or, l'automate complémentaire de $\mathcal{A} = (Q, \Sigma, q_i, \delta, F)$ est l'automate fini $\overline{\mathcal{A}} = (Q, \Sigma, q_i, \delta, Q \setminus F)$ et $\overline{\mathcal{A}}$ reconnaît l'ensemble des mots non reconnus par $\mathcal{A}$. Donc, $\overline{\mathcal{L}}$ est un langage reconnaissable et donc régulier.

C2. Soit Σ un alphabet, $\mathcal{L}_1$ et $\mathcal{L}_2$ deux langages réguliers sur Σ. Montrer que $\mathcal{L}_1 \cap \mathcal{L}_2$ est un langage régulier.

> **Solution :** On passe par la loi de Morgan : $\mathcal{L}_1 \cap \mathcal{L}_2 = \overline{\overline{\mathcal{L}_1} \cup \overline{\mathcal{L}_2}}$. Or, les langages réguliers sont stables pour l'union (par construction) et la complémentation (on vient le montrer). Donc, les langages réguliers sont stables pour l'intersection.

C3. Soit Σ un alphabet, $\mathcal{L}_1$ et $\mathcal{L}_2$ deux langages réguliers sur Σ. Montrer que $\mathcal{L}_1 \setminus \mathcal{L}_2$ est un langage régulier.

> **Solution :** On a $\mathcal{L}_1 \setminus \mathcal{L}_2 = \mathcal{L}_1 \cap \overline{\mathcal{L}_2}$. Comme les langages réguliers sont stables pour l'intersection et la complémentation, ils sont également stables pour la différence ensembliste.

C4. Soit Σ un alphabet et $\mathcal{L}$ un langage régulier sur Σ. On définit le mot miroir de $w = a_1 a_2 \dots a_n$ par $w^R = a_n a_{n-1} \dots a_1$ et le langage miroir $\mathcal{L}^R = \{u \in \Sigma^*, u^R \in \mathcal{L}\}$. Montrer que $\mathcal{L}^R$ est régulier.

> **Solution :** On considère un automate fini non déterministe $\mathcal{A} = (Q, \Sigma, Q_i, \Delta, F)$ associé à $\mathcal{L}$. Soit l'automate fini $\mathcal{A}^R = (Q, \Sigma, F, \Delta^{-1}, Q_i)$. On a noté Δ^{-1} l'ensemble des transitions obtenues en inversant le sens de chaque transition de Δ : par exemple, (q, a, q') devient (q', a, q). $\mathcal{A}^R$ reconnaît le langage miroir $\mathcal{L}^R$, car pour tout mot w de $\mathcal{L}$ il existe un chemin acceptant dans $\mathcal{A}$ qui correspond à un chemin acceptant w^R dans $\mathcal{A}^R$. $\mathcal{L}^R$ est un langage reconnaissable donc régulier. Les langages réguliers sont stables par miroir.

C5. Soit Σ un alphabet et $\mathcal{L}$ un langage régulier sur Σ. Montrer que $\mathrm{Pref}(\mathcal{L}) = \{u \in \Sigma^*, \exists v \in \Sigma^*, uv \in \mathcal{L}\}$, l'ensemble de mots préfixes de $\mathcal{L}$, est un langage régulier.

> **Solution :** Soit $\mathcal{A} = (Q, \Sigma, q_i, \delta, F)$ un automate fini déterministe reconnaissant $\mathcal{L}$. Soit C l'ensemble des états co-accessibles de $\mathcal{A}$. Alors l'automate $\mathcal{A}_{pref} = (Q, \Sigma, q_i, \delta, C)$ est un automate fini qui reconnaît $\mathrm{Pref}(\mathcal{L})$. Les langages réguliers sont donc stables pour l'opération préfixe.

C6. Soit Σ un alphabet et $\mathcal{L}$ un langage régulier sur Σ. Montrer que $\mathrm{Suff}(\mathcal{L}) = \{v \in \Sigma^*, \exists u \in \Sigma^*, uv \in \mathcal{L}\}$, l'ensemble de mots suffixes de $\mathcal{L}$, est un langage régulier.

> **Solution :** Soit $\mathcal{A} = (Q, \Sigma, q_i, \delta, F)$ un automate fini déterministe reconnaissant $\mathcal{L}$. Soit A l'ensemble des états accessibles de $\mathcal{A}$. Alors l'automate $\mathcal{A}_{pref} = (Q, \Sigma, A, \delta, F)$ est un automate fini non déterministe [1] qui reconnaît $\mathrm{Suff}(\mathcal{L})$. Les langages réguliers sont donc stables pour l'opération suffixe.

C7. Soit Σ un alphabet et $\mathcal{L}$ un langage régulier sur Σ. Montrer que $\mathrm{Fact}(\mathcal{L}) = \{w \in \Sigma^*, \exists u, v \in \Sigma^*, uwv \in \mathcal{L}\}$, l'ensemble de mots facteurs de $\mathcal{L}$, est un langage régulier.

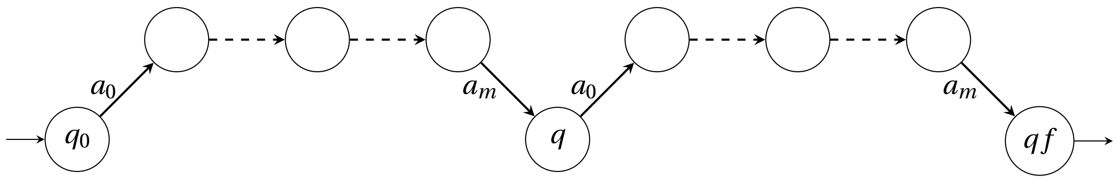
> **Solution :** On remarque que $\mathrm{Fact}(\mathcal{L}) = \mathrm{Pref}(\mathrm{Suff}(\mathcal{L}))$. Les langages réguliers sont donc stables pour l'opération facteur.

C8. Soit Σ un alphabet et $\mathcal{L}$ un langage régulier sur Σ. On définit le langage racine $\sqrt{\mathcal{L}}$ par :

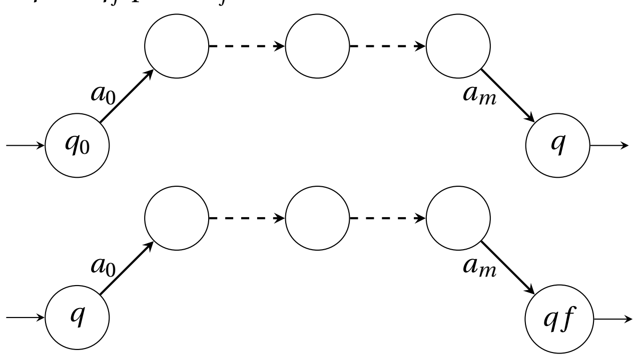
$$\sqrt{\mathcal{L}} = \{u \in \Sigma^*, uu \in \mathcal{L}\} \tag{1}$$

Montrer que $\sqrt{\mathcal{L}}$ est régulier en vous appuyant sur l'automate reconnaissant le langage $\mathcal{L}$.

Solution : Soit $\mathcal{A} = (Q, \Sigma, q_i, \delta, F)$ un automate fini déterministe complet reconnaissant le langage régulier $\mathcal{L}$. Soit $w = a_0 a_1 \dots a_m$ un mot de $\sqrt{\mathcal{L}}$. Nécessairement, le mot ww est reconnu par $\mathcal{A}$. Cela signifie qu'il existe un état q et un chemin dans l'automate tel que :



Dire que w est un mot de $\sqrt{\mathcal{L}}$ est donc équivalent à dire que ce mot est reconnu à la fois par l'automate $\mathcal{A}_d = (\{q_0, \dots, q\}, \Sigma, q_0, \delta, \{q\})$ et par l'automate $\mathcal{A}_f = (\{q, \dots, q_f\}, \Sigma, q, \delta, \{q_f\})$. Les ensembles des états sont obtenus en sélectionnant les états du chemin $q_0 \longrightarrow q$ pour $\mathcal{A}_d$ et ceux de $q \longrightarrow q_f$ pour $\mathcal{A}_f$.



Essayer de construire un automate qui calcule ces chemins en parallèle. Supposons que l'automate $\mathcal{A}$ possède n états distincts : $Q = q_0, q_1, \dots, q_n$ et considérons l'automate :

$$\mathcal{A}_p = (Q^n, \Sigma, (q_0, q_1, \dots, q_n), \delta_p, F_p) \tag{2}$$

La fonction de transition δ_p est construire ainsi :

$$\delta((p_0, p_1, \dots, p_n), a) = (\delta(p_0, a), \delta(p_1, a), \dots, \delta(p_n, a)) \tag{3}$$

L'ensemble des états accepteurs F_p est défini comme suit :

$$F_p = \{(p_0, p_1, \dots, p_n), p_i \in F \textbf{ et } p_0 = q_i\} \tag{4}$$

$\mathcal{A}_p$ est un automate fini (car Q^n est de cardinal n^n) déterministe (car δ est une fonction et donc δ_p également) qui reconnait $\sqrt{\mathcal{L}}$ (car l'ensemble des états accepteurs vérifient $p_i \in F$ **et** $p_0 = q_i$). $\sqrt{\mathcal{L}}$ est donc un langage régulier.

C9. Montrer que tout langage fini est un langage régulier.

Solution : Soit $\mathcal{L}$ un langage fini de cardinal n. Chaque mot w_i de ce langage fini peut être représenté par une expression régulière e_i, celle qui utilise toutes les lettres nécessaires de l'alphabet pour le représenter. D'après la sémantique des expressions régulières, on a :

$$\mathcal{L} = \bigcup_{i \in [\![0, n-1]\!]} \mathcal{L}_{ER}(e_i) = \mathcal{L}_{ER}(e_0|e_1|\dots|e_{n-1})$$

Par définition, cela signifie que $\mathcal{L}$ est un langage dénoté par une expression régulière. Par conséquent, un langage fini est toujours régulier.

D Lemme de l'étoile et non régularité

D1. Montrer que les langages ci-dessous ne sont pas réguliers.

(a) $\mathcal{L}_1 = \{a^p, \text{p est premier}\}$

Solution : Supposons que $\mathcal{L}_1$ soit régulier. Soit une constante d'itération n de ce langage et un mot w de cardinal supérieur ou égal n.

Soit p un nombre premier strictement plus grand que n et $w = a^p$, un mot de le mot de $\mathcal{L}_1$. D'après le lemme de l'étoile, on peut décomposer w en xyz tels que $|xy| \leqslant n$, $y \neq \epsilon$ et $xy^*z \subseteq \mathcal{L}_1$.

Soit k un entier strictement inférieur à n. On pose $x = a^{k-1}$, $y = a$ et $z = a^{p-k}$. Une telle décomposition satisfait les conditions du lemme de l'étoile. On peut donc itérer sur y. En particulier, on peut choisir d'élever y à la puissance $p+1$ et obtenir u, un mot de $\mathcal{L}_1$. $u = xy^{p+1}z = a^{k-1}a^{p+1}a^{p-k} = a^{2p}$. Or, si $2p$ n'est pas premier et donc $u \notin \mathcal{L}_1$. On aboutit donc à une contradiction. $\mathcal{L}_1$ n'est pas un langage régulier.

(b) $\mathcal{L}_2 = \{w \in \Sigma^*, \Sigma = \{a, b\}, w \text{ possède autant de a que de b}\}$

Solution : Supposons que $\mathcal{L}_2$ soit régulier. Soit une constante d'itération n de ce langage et un mot de $\mathcal{L}_2$ de cardinal supérieur ou égal n. On pose par exemple $w = a^n b^n$ dont le cardinal vaut $2n$. On peut appliquer le lemme de l'étoile à w.

Soit k un entier strictement inférieur à n. On pose $x = a^{k-1}$, $y = a$ et $z = a^{n-k}b^n$. Une telle décomposition satisfait les conditions du lemme de l'étoile et on peut donc itérer sur y. En particulier, on peut élever y à la puissance zéro et obtenir un mot u de $\mathcal{L}_2$: $u = xy^0z = a^{k-1}a^{n-k}b^n = a^{n-1}b^n$. Or, le nombre de a de u n'est pas égal au nombre de b de u. Donc $u \notin \mathcal{L}_2$. On aboutit donc à une contradiction. $\mathcal{L}_2$ n'est pas un langage régulier.

(c) $\mathcal{L}_3 = \{a^i b^j, i < j\}$

Solution : Supposons que $\mathcal{L}_3$ soit régulier. Soit une constante d'itération n de ce langage et un mot de $\mathcal{L}_3$ de cardinal supérieur ou égal n. On pose par exemple $w = a^n b^{n+1}$ dont le cardinal vaut $2n+1$. On peut appliquer le lemme de l'étoile à w.

Soit k un entier strictement inférieur à n. On pose $x = a^{k-1}$, $y = a$ et $z = a^{n-k}b^{n+1}$. Une telle décomposition satisfait les conditions du lemme de l'étoile et on peut donc itérer sur y. En particulier, on peut élever y à la puissance trois et obtenir un mot u de $\mathcal{L}_3$: $u = xy^3z = a^{k-1}a^3a^{n-k}b^{n+1} = a^{n+2}b^n$. Or, dans ce cas, comme $n+2 > n$, $u \notin \mathcal{L}_3$. On aboutit donc à une contradiction. $\mathcal{L}_3$ n'est pas un langage régulier.

(d) $\mathcal{L}_4 = \{a^p, \text{p n'est pas premier}\}$

Solution : Comme les langages réguliers sont stables par complémentation et que $\mathcal{L}_1$ n'est pas régulier, $\mathcal{L}_4$ n'est pas régulier.

D2. Montrer que l'ensemble des mots de Dyck $\mathcal{D}$ n'est pas un langage régulier. On rappelle que $\mathcal{D}$ est l'ensemble de mots bien parenthésés sur un alphabet fini de parenthèses ouvrantes et fermantes. Par exemple, sur la paire de parenthèses formée de (et), le mot (())() est un mot bien parenthésé, alors que le mot ())(ne l'est pas.

Solution : En notant les parenthèses ouvrantes et fermantes a et b, dire que l'ensemble des mots de Dyck est un langage régulier implique que l'ensemble $\{a^nb^n, n \in \mathbb{N}\} \subset \mathcal{D}$ est un langage régulier. Or ce n'est pas le cas car il s'agit du langage des puissances qui n'est pas régulier (cf. cours).

Plus précisément, on peut écrire : $\{a^nb^n, n \in \mathbb{N}\} = \mathcal{D} \cap \mathcal{L}_{ER}(a^*b^*)$. Or, l'intersection de deux langages réguliers est un langage régulier. Donc si $\mathcal{D}$ était régulier, le langage des puissances le serait aussi puisque $\mathcal{L}_{ER}(a^*b^*)$ est un langage régulier car dénoté par une expression régulière.

D3. Soit $\Sigma = \{a, b\}$ un alphabet à deux lettres. et $\mathcal{L}_{pal}$ lensemble des palindromes sur Σ. Montrer que $\mathcal{L}_{pal}$ nest pas un langage régulier.

Solution : Supposons que $\mathcal{L}_{pal}$ soit régulier et n une constante d'itération de ce langage régulier.

Soit w un mot de $\mathcal{L}_{pal}$ de longueur supérieure ou égale à n. Considérons le mot $w = a^nba^n$ qui est un palindrome. Ce mot appartient au langage $\mathcal{L}_i$ intersection de $\mathcal{L}_{pal}$ et de $\mathcal{L}_{ER}(a^*ba^*)$. Comme $\mathcal{L}_{ER}(a^*ba^*)$ est un langage régulier et que l'intersection de deux langages réguliers est un langage régulier, si $\mathcal{L}_{pal}$ est régulier alors $\mathcal{L}_i$ l'est aussi.

On a $|w| = 2n+1 > n$ et on peut appliquer le lemme de l'étoile à $w \in \mathcal{L}_i$. On peut décomposer w en xyz tels que $|xy| = k \leqslant n$ et $y \neq \epsilon$. Selon cette décomposition, il existe donc deux entiers i et $j > 0$ tels que $i + j = k$, $x = a^i$, $y = a^j \neq \epsilon$ et $z = a^{n-i-j}ba^n$. Le lemme de l'étoile nous dit que l'on peut itérer sur y et obtenir un mot de $\mathcal{L}_i$, $xy^*z \in \mathcal{L}_i$.

Soit le mot u obtenu en prenant la puissance nulle de y : $u = xy^0z = a^ia^{n-i-j}ba^n = a^{n-j}ba^n$. Comme $j > 0$, ce mot n'est pas un palindrome et donc n'appartient pas au langage $\mathcal{L}_i$, ce qui est en contradiction avec le lemme de l'étoile : $\mathcal{L}_i$ n'est pas régulier. Donc $\mathcal{L}_{pal}$ n'est pas un langage régulier.