Langages formels

Exercices 01

Partie 1 : Mots

Exercice 1

Soit $\mathcal{A}$ un alphabet : on suppose que $\mathcal{A}$ est muni d'un ordre total noté $\leqslant$.

1. Montrer qu'il existe une bijection $\phi : \mathcal{A} \to [\![0, |\mathcal{A}|-1]\!]$ telle que pour tout $a, b \in \mathcal{A}$, $a \leqslant b \iff \phi(a) \leqslant \phi(b)$ (en moins formel : tout ordre total sur un ensemble fini est assimilable à un ordre linéaire).

On pose $\leqslant_{\text{lex}}$ l'ordre lexicographique sur $\mathcal{A}^*$: pour m et m' deux mots, on a $m \leqslant m'$ ssi, en notant p le préfixe commun le plus long de m et m', avec $m = pq$ et $m' = pq'$, $q = \varepsilon$ ou $q_0 \leqslant q'_0$.

2. Exemple : soit $\mathcal{B} = \{a, b\}$, et $\leqslant$ l'ordre alphabétique latin usuel. Déterminer si $abba \leqslant abab$ et si $abba \leqslant abbabaab$.
3. Montrer que l'ordre lexicographique est bien un ordre sur $\mathcal{A}^*$; montrer qu'il est total, mais pas bien fondé.

Exercice 2

Soit $\mathcal{A}$ un alphabet et m un mot. La complexité de m est la fonction $p_m : [\![0, |m|]\!] \to \mathbb{N}$ qui, à un entier k, associe le nombre de facteurs distincts de m de longueur k.

1. Soit $m = banana$. Déterminer $p_m(k)$ pour $k \in [\![0, 6]\!]$.
2. Montrer que pour tout k, $p_m(k) \leqslant |\mathcal{A}|^k$.
3. Montrer qu'en fait, pour tout k, $p_m(k) \leqslant \min(|\mathcal{A}|^k, |m| - k + 1)$.

On peut étendre la notion de complexité aux mots infinis : un mot infini est un élément de $\mathcal{A}^{\mathbb{N}}$, et la complexité d'un mot infini m est définie sur $\mathbb{N}$.

4. Soit $m = aaaaaa\ldots$ Déterminer $p_m(k)$ pour tout $k \in \mathbb{N}$.
5. Même question pour $m = abbaaaaaaaa\ldots$
6. Montrer que pour tout k, $p_m(k) \leqslant |\mathcal{A}|^k$.
7. Exhiber un mot infini tel que pour tout $k \in \mathbb{N}$, $p_m(k) = |\mathcal{A}|^k$.

Exercice 3

Soit $n \geqslant 1$, soit $\mathcal{A}$ un alphabet **binaire**. On pose $G_n = (\mathcal{A}^n, V)$ le graphe orienté étiqueté des mots de longueur n tel que $m \xrightarrow{a} m'$ ssi $m' \sqsupseteq_s ma$.

1. Déterminer G_3.
2. Montrer que pour tout mot m, $\deg_S(m) = 2$.
3. En déduire qu'il existe un cycle eulérien dans G_n.
4. Un mot de **de Bruijn** est un mot contenant comme facteur de tout élément de $\mathcal{A}^n$ avec une seule occurrence. Montrer que pour tout $n \in \mathbb{N}^*$, un tel mot existe.
5. On suppose désormais que l'alphabet est ternaire : existe-t-il un mot de de Bruijn sur un tel alphabet ?

Exercice 4

Avec l'alphabet $\mathcal{A} = \{(,)\}$, on dit qu'un mot est **bien parenthésé** s'il vérifie les propriétés suivantes :

— ε est bien parenthésé ;

— si m est bien parenthésé, alors (m) est aussi bien parenthésé ;
— si m et m' sont bien parenthésés, alors $m \cdot m'$ est aussi bien parenthésé.

1. Déterminer si $((()())())$, $)(())$ et $())(()$ sont bien parenthésés.
2. Donner un algorithme de complexité linéaire en la longueur du mot décidant si un mot est bien parenthésé.
3. On note C_n le nombre de mots bien parenthésés de longueur $2n$. Montrer que :

$$C_{n+1} = \sum_{k=0}^{n} C_k C_{n-k}$$

4. Donner un algo de complexité polynomiale donnant le nombre de mots bien parenthésés de longueur $2n$.

Exercice 5

1. Un mot fini est dit *primitif* s'il ne peut pas s'écrire sous forme d'une puissance non-nulle. Montrer que tout mot x peut s'écrire de manière unique sous la forme r^p, où r est un mot primitif et $p \in \mathbb{N}$ (r est appelé la racine primitive de x).
2. Soient x et y deux mots finis tels que $xy = yx$. Montrer qu'il existe $w \in \mathcal{A}^*$ et deux entiers p et $q \in \mathbb{N}$ tels que $x = w^p$ et $y = w^q$.
3. En déduire, pour tout mot fini x, le commutant de x : $\mathrm{Com}(x) = \{y \in \mathcal{A}^* \mid xy = yx\}$.

Exercice 6

Soit $\mathcal{A} = \{0, 1\}$. On dit qu'un mot fini x est équilibré si pour toute paire (f, g) de facteurs de x vérifiant $|f| = |g|$, on a $||f|_0 - |g|_0| \leqslant 1$.

1. Soit $x = 0101$. Montrer que x est équilibré.
2. Soit $x = 0010010101$. Montrer que x n'est pas équilibré.
3. Montrer qu'un mot x n'est pas équilibré ssi il existe $m \in \mathcal{A}^*$ tel que $0m0$ et $1m1$ sont des facteurs de x.

Partie 2 : Langages rationnels

Exercice 7

Pour un langage $L \subseteq \mathcal{A}^*$, on note les opérations suivantes :

— $L^+ = \{m_1 \cdot m_2 \cdot \ldots m_k \mid k \geqslant 1, \forall i \in [\![1, k]\!], m_i \in L\}$;
— $L? = L \cup \{\varepsilon\}$.

Montrer que les langages rationnels sont stables par ces deux opérations.

Exercice 8

Soient $u, v \in \mathcal{A}^*$. Montrer qu'il existe un unique $w \in \mathcal{A}^*$ tel que $u^* \cap v^* = w^*$.

Exercice 9 : lemme d'Arden

Soient A et $B \subseteq \mathcal{A}^*$ deux langages (pas forcément rationnels). On considère l'équation suivante, d'inconnue $L \subseteq \mathcal{A}^*$:

$$L = AL + B$$

1. Montrer que $L_s = A^*B$ est solution de l'équation.

2. Supposons de plus que $\varepsilon \notin A$: montrer que L_s est l'unique solution de l'équation.
3. Exhiber un exemple où $\varepsilon \in A$ et l'équation a plusieurs solutions.

Exercice 10

Soit $L \subseteq \mathcal{A}^*$ un langage. On dit qu'un langage $M \subseteq \mathcal{A}^*$ vérifie la propriété P_L s'il vérifie les trois conditions suivantes :

— $\varepsilon \in M$;

— $L \subseteq M$;

— $M^2 \subseteq M$.

Montrer que L^* est le plus petit langage vérifiant la propriété P_L.

Exercice 11

Pour un langage $L \in \mathcal{A}^*$, on note $\widetilde{L}$ le miroir de L :

$$\widetilde{L} = \{m_{|m|-1}m_{|m|-2}\dots m_1m_0 \mid m_0m_1\dots m_{|m|-2}m_{|m|-1} \in L\}$$

Montrer que $\mathrm{Rat}(\mathcal{A}^*)$ est stable par miroir.

Exercice 12

Pour un langage $L \in \mathcal{A}^*$, on note $\mathrm{Pref}(L)$ l'ensemble des préfixes de L :

$$\mathrm{Pref}(L) = \{m \mid \exists x \in L, m \sqsubseteq_p x\}$$

Montrer que $\mathrm{Rat}(\mathcal{A}^*)$ est stable par préfixation.

Partie 3 : Automates

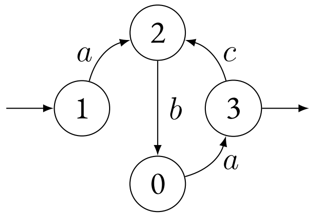
Exercice 13

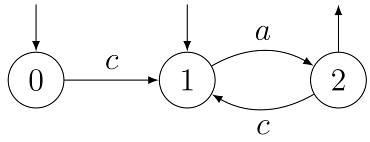
Soit $\mathcal{A} = \{a, b\}$. Déterminer si les langages suivants sont rationnels :

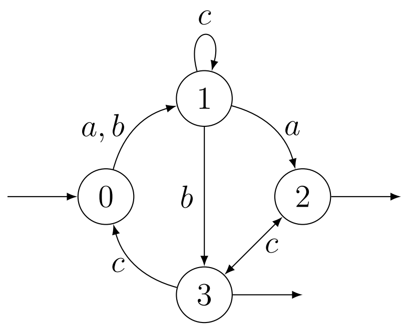
1. $\{a^p \mid p \text{ premier }\}$;
2. $\{m \in \mathcal{A}^* \mid |m|_a = |m|_b\}$;
3. $\{a^{2n}b^n \mid n \in \mathbb{N}\}$;
4. $\{a^nb^m \mid n \neq m\}$;
5. $\{a^{2^k} \mid k \in \mathbb{N}\}$;
6. $\{a^{k^2} \mid k \in \mathbb{N}\}$;
7. $\{m \cdot m \mid m \in \mathcal{A}^*\}$;
8. $\{\text{palindromes}\}$;
9. $\{a^{2^k} \mid k \in \mathbb{N}\}$;
10. $\{\text{codage binaire des nombres premiers}\}$.

Exercice 14

Appliquer l'algorithme d'élimination d'états aux automates suivants, en éliminant les états dans l'ordre croissant :







Exercice 15

Pour obtenir une expression régulière à partir d'un automate fini, un autre algorithme est l'algorithme de McNaughton-Yamada, dont l'idée est la suivante :
Soit $\mathfrak{A} = (\mathcal{A}, Q, I, F, T)$ un automate fini. Quitte à renomer les états, on suppose que $Q = [\![0, N]\!]$. Pour toute paire d'états p et q, on définit alors $L_{p,q} = \{m \in \mathcal{A}^* \mid p \xrightarrow{m}^* q \text{ dans } \mathfrak{A}\}$.

1. Décrire $\mathcal{L}(\mathfrak{A})$ à partir des $L_{p,q}$.
2. On définit $L_{p,q}^{(k)} = \{m \in * \mid p \xrightarrow{m}^* q \text{ dans } \mathfrak{A} \text{ où tout état intermédiaire est } < k\}$.
   a) Décrire $L_{p,q}$ à l'aide des $L_{p,q}^{(k)}$.
   b) Déterminer, pour toute paire (p, q), une expression régulière pour $L_{p,q}^{(0)}$.
   c) Soit k fixé. On suppose disposer d'expressions régulières pour tout $L_{p,q}^{(k)}$. Donner une méthode permettant d'obtenir une expression régulière pour $L_{p,q}^{(k+1)}$ pour une paire d'états (p, q).
   d) Donner un pseudo-code implémentant l'algorithme de McNaughton-Yamada.
3. Déterminer la complexité temporelle de l'algorithme de McNaughton-Yamada.
4. Appliquer l'algorithme de McNaughton-Yamada sur l'automate fini suivant :

Exercice 16

Pour un langage $L \in \mathcal{A}^*$, on note $\sqrt{L}$ la racine de L :

$$\sqrt{L} = \{u \in \mathcal{A}^* \mid u \cdot u \in L\}$$

Montrer que $\text{Rat}(\mathcal{A}^*)$ est stable par passage à la racine.

Exercice 17 : Centrale 2023

Pour deux langages L et $K \subseteq \mathcal{A}^*$, on note $L \rhd K$ le langage suivant :

$$L \rhd K = \{uwv \mid uv \in L, w \in K\}$$

Montrer que $\text{Rat}(\mathcal{A}^*)$ est stable par l'opération $\rhd$.

Exercice 18

Pour deux mots finis x et y, le **shuffle** (ou mélange) de x et y est le langage :

$$x \diamond y = \{x_0 y_0 x_1 y_1 \dots x_n y_n \mid \forall i, x_i, y_i \in \mathcal{A}^*, x = x_0 x_1 \dots x_n \text{ et } y = y_0 y_1 \dots y_n\}$$

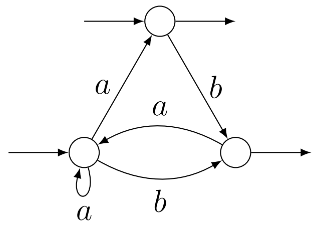
Attention : les x_i et y_i sont des mots, potentiellement vides.

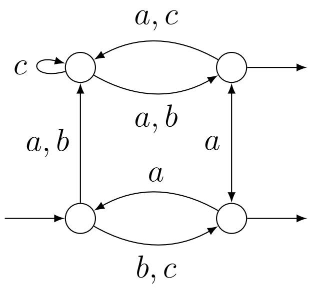
1. Soient $x = ab$ et $y = cde$. Déterminer $x \diamond y$.
2. Pour deux langages X et Y, on définit $X \diamond Y = \bigcup\limits_{x \in X, y \in Y} x \diamond y$. Montrer que si X et Y sont réguliers, alors

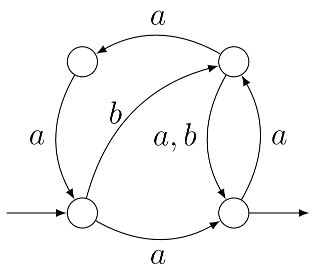
$X \diamond Y$ est régulier.

Exercice 19

Déterminiser les automates suivants :







Exercice 20 : résiduels

Pour un langage $L \subseteq \mathcal{A}^*$ et un mot $x \in \mathcal{A}^*$, on définit le langage :

$$x^{-1}L = \{y \in \mathcal{A}^* \mid xy \in L\}$$

Pour deux mots x et $y \in \mathcal{A}^*$, on note $x \sim y$ ssi $x^{-1}L = y^{-1}L$.

1. Montrer que $\sim$ est bien une relation d'équivalence.
2. Montrer que pour $x, y \in \mathcal{A}^*$ et $a \in \mathcal{A}$, si $x \sim y$, alors $x \cdot a \sim y \cdot a$.
3. Montrer que si $\sim$ admet un nombre fini de classes d'équivalence, alors L est régulier.
4. Montrer que si L est régulier, alors $\sim$ admet un nombre fini de classes d'équivalence.
   Indication : si L est régulier, alors L est reconnu par un automate déterministe complet.
5. (*) Montrer qu'il existe, à isomorphisme près, un unique automate déterministe minimal reconnaissant L.