TD 2 – Automates finis et expressions rationnelles

1 Langages locaux

Soit Σ un alphabet. On dit qu'un langage L sur Σ est *local* lorsqu'il existe $P, S \subseteq \Sigma$ et $N \subseteq \Sigma^2$ tels que $L \setminus \{\varepsilon\} = (P\Sigma^* \cap \Sigma^* S) \setminus \Sigma^* N \Sigma^*$.

Exercice 1.

Caractériser la classe des langages locaux en termes d'AFD, c'est-à-dire, définir une classe d'AFD dits *locaux* telle qu'un langage est local si et seulement s'il est reconnu par un AFD local.

Solution 1.

C'est la classe des AFD $(Q, \Sigma, \delta, q_0, F)$ tels qu'il existe $p \in Q \setminus F$ vérifiant $\delta(p, a) = p$ pour tout $a \in \Sigma$ (p est un puits non acceptant) et que pour tout $a \in \Sigma$, on ait $|\{\delta(q, a) \mid q \in Q\} \setminus \{p\}| \leq 1$ (pour toute lettre lue, quel que soit l'état dans lequel on se trouve, on va soit dans le puits p, soit dans un autre état qui est toujours le même).

Soit $\mathcal{A} = (Q, \Sigma, \delta, i, F)$ un tel AFD. Pour tout $a \in \Sigma$, on pose q_a comme étant l'unique état de $\{\delta(q, a) \mid q \in Q\} \setminus \{p\}$ si celui-ci existe, p sinon. Posons :

— $P = \{a \in \Sigma \mid \delta(i, a) \neq p\}$;

— $S = \{a \in \Sigma \mid q_a \in F\}$;

— $N = \{a_1 a_2 \in \Sigma^2 \mid \delta(q_{a_1}, a_2) = p\}$.

Il est direct de voir que, pour tous $q \in Q$ et $a \in \Sigma$, on a $\delta(q, a) \in \{q_a, p\}$. Ainsi donc, pour tout mot $w = a_1 \cdots a_n$ avec $n \in \mathbb{N}_{>0}$ et $a_1, \ldots, a_n \in \Sigma$, on a $\hat{\delta}(i, w) \in \{q_{a_n}, p\}$. En utilisant cette observation, on peut montrer que $\mathcal{L}(\mathcal{A}) \setminus \{\varepsilon\} = (P\Sigma^* \cap \Sigma^* S) \setminus \Sigma^* N \Sigma^*$.

Soit maintenant L un langage local sur un alphabet Σ. Il existe donc $P, S \subseteq \Sigma$ et $N \subseteq \Sigma^2$ tels que $L \setminus \{\varepsilon\} = (P\Sigma^* \cap \Sigma^* S) \setminus \Sigma^* N \Sigma^*$. Construisons l'automate $\mathcal{A} = (\Sigma \cup \{q_0, p\}, \Sigma, \delta, q_0, F)$ où $\{q_0, p\} \cap \Sigma = \emptyset$, $F = \begin{cases} S \cup \{q_0\} & \text{si } \varepsilon \in L \\ S & \text{sinon} \end{cases}$ et $\delta \colon (\Sigma \cup \{q_0, p\}) \times \Sigma \to (\Sigma \cup \{q_0, p\})$ est telle que :

— pour tout $a \in \Sigma$, $\delta(q_0, a) = \begin{cases} a & \text{si } a \in P \\ p & \text{sinon} \end{cases}$;

— pour tous $a_1, a_2 \in \Sigma$, $\delta(a_1, a_2) = \begin{cases} a_2 & \text{si } a_1 a_2 \notin N \\ p & \text{sinon} \end{cases}$;

— pour tout $a \in \Sigma$, $\delta(p, a) = p$.

On peut montrer que $\mathcal{A}$ vérifie les conditions énoncées plus haut et que $L = \mathcal{L}(\mathcal{A})$.

Exercice 2.

Soit L_1 un langage local sur un alphabet Σ_1 et L_2 un langage local sur un alphabet Σ_2 tels que $\Sigma_1 \cap \Sigma_2 = \emptyset$. Montrer que $L_1 \cup L_2$ et $L_1 L_2$ sont locaux.

Solution 2.

La démonstration se fait assez directement en utilisant la caractérisation de l'exercice précédent.

Exercice 3.

Soit L un langage local sur un alphabet Σ. Montrer que L^* est local.

Solution 3.

La démonstration se fait également assez directement en utilisant la caractérisation de l'exercice 1.

2 Algorithme de Glushkov

Une expression rationnelle sur un alphabet Σ est dite *linéaire* lorsqu'elle contient au plus une fois chaque lettre $a \in \Sigma$.

Exercice 4.

Montrer que toute expression rationnelle linéaire correspond à un langage local.

Solution 4.

On procède simplement par récurrence sur la forme d'une expression rationnelle linéaire en utilisant les deux résultats précédents.

Exercice 5.

Étant donné une expression rationnelle linéaire E sur un alphabet Σ, donner un algorithme permettant de déterminer $P, S \subseteq \Sigma$ et $N \subseteq \Sigma^2$ tels que $\mathcal{L}(E) \setminus \{\varepsilon\} = (P\Sigma^* \cap \Sigma^* S) \setminus \Sigma^* N \Sigma^*$, ainsi que si $\varepsilon \in \mathcal{L}(E)$ ou non.

Solution 5.

Étant donné une expression rationnelle linéaire E, on calcule récursivement sur la forme de E les ensembles $P(E) \subseteq \Sigma$, $S(E) \subseteq \Sigma$, $\epsilon(E) \subseteq \{\varepsilon\}$ et $F(E) \subseteq \Sigma^2$ tels que $\mathcal{L}(E) \setminus \{\varepsilon\} = (P(E)\Sigma^* \cap \Sigma^* S(E)) \setminus \Sigma^*(\Sigma^2 \setminus F(E))\Sigma^*$ et $\varepsilon \in \mathcal{L}(E)$ si et seulement si $\varepsilon \in \epsilon(E)$.

On procède de la manière suivante, pour tous $a \in \Sigma$ et E_1, E_2 expressions rationnelles linéaires.

$$\begin{aligned}
P(\emptyset) &= \emptyset \\
P(\varepsilon) &= \emptyset \\
P(a) &= \{a\} \\
P(E_1 + E_2) &= P(E_1) \cup P(E_2) \\
P(E_1 E_2) &= P(E_1) \cup \epsilon(E_1) P(E_2) \\
P(E_1^*) &= P(E_1) \ ,
\end{aligned}$$

$$\begin{aligned}
S(\emptyset) &= \emptyset \\
S(\varepsilon) &= \emptyset \\
S(a) &= \{a\} \\
S(E_1 + E_2) &= S(E_1) \cup S(E_2) \\
S(E_1 E_2) &= S(E_2) \cup S(E_1) \epsilon(E_2) \\
S(E_1^*) &= S(E_1) \ ,
\end{aligned}$$

$$\begin{aligned}
\epsilon(\emptyset) &= \emptyset \\
\epsilon(\varepsilon) &= \{\varepsilon\} \\
\epsilon(a) &= \emptyset \\
\epsilon(E_1 + E_2) &= \epsilon(E_1) \cup \epsilon(E_2) \\
\epsilon(E_1 E_2) &= \epsilon(E_1) \epsilon(E_2) \\
\epsilon(E_1^*) &= \{\varepsilon\} \ ,
\end{aligned}$$

$$\begin{aligned}
F(\emptyset) &= \emptyset \\
F(\varepsilon) &= \emptyset \\
F(a) &= \emptyset \\
F(E_1 + E_2) &= F(E_1) \cup F(E_2) \\
F(E_1 E_2) &= F(E_1) \cup F(E_2) \cup S(E_1) P(E_2) \\
F(E_1^*) &= F(E_1) \cup S(E_1) P(E_1) \ .
\end{aligned}$$

Exercice 6.

En déduire un algorithme permettant, étant donné une expression rationnelle E, de construire un AFN reconnaissant $\mathcal{L}(E)$.

Solution 6.

Il suffit, étant donné l'expression rationnelle E, de la linéariser en une expression rationnelle linéaire E' en mémorisant les remplacements de lettres effectués, de calculer $P(E')$, $S(E')$, $\epsilon(E')$ et $F(E')$ en utilisant la procédure récursive de l'exercice précédent, de construire l'AFD correspondant à $\mathcal{L}(E')$ en utilisant la construction utilisée dans la solution de l'exercice 1, puis d'y remplacer les lettres dans le sens inverse, ce qui donne un AFN reconnaissant $\mathcal{L}(E)$.

Exercice 7.

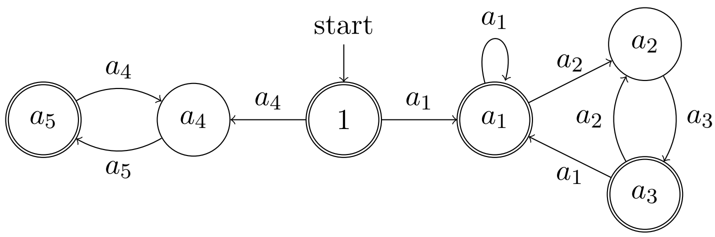
Utiliser cet algorithme pour obtenir un AFN reconnaissant le langage correspondant à $(a(ab)^*)^* + (ba)^*$ sur $\{a, b\}$.

Solution 7.

L'expression rationnelle linéarisée est $E' = (a_1(a_2a_3)^*)^* + (a_4a_5)^*$.

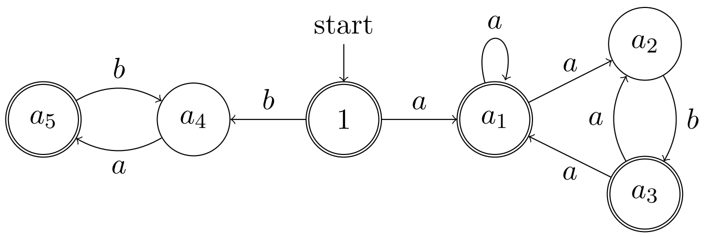
En calculant à la main, on obtient $P(E') = \{a_1, a_4\}$, $S(E') = \{a_1, a_3, a_5\}$, $\epsilon(E') = \{\varepsilon\}$, $F(E') = \{a_1a_1, a_1a_2, a_2a_3, a_3a_1, a_3a_2, a_4a_5, a_5a_4\}$.

On construit l'AFD suivant reconnaissant $\mathcal{L}(E')$ (où on omet l'état puits) :



.

On le transforme enfin en l'AFN suivant reconnaissant $\mathcal{L}((a(ab)^*)^* + (ba)^*)$:



.

3 Équations linéaires

Exercice 8. *Lemme d'Arden*

Soit K et L deux langages sur un alphabet Σ. Montrer que si $\varepsilon \notin K$, alors il existe un unique $X \subseteq \Sigma^*$ tel que $X = KX \cup L$ en explicitant X.

Solution 8.

Supposons que $\varepsilon \notin K$ et montrons que l'unique $X \subseteq \Sigma^*$ tel que $X = KX \cup L$ est $X = K^*L$.

Soit $X \subseteq \Sigma^*$ tel que $X = KX \cup L$. Nous allons montrer que pour tout $w \in \Sigma^*$, $w \in X$ si et seulement si $w \in K^*L$, par récurrence sur la longueur de w.

Cas de base $|w| = 0$. Si $\varepsilon \in K^*L$, alors $\varepsilon \in L$ et donc $\varepsilon \in X$. Inversement, si $\varepsilon \in X$, alors, puisque $\varepsilon \notin K$, on a $\varepsilon \in L$ et donc $\varepsilon \in K^*L$.

Récurrence. Soit $l \in \mathbb{N}_{>0}$ et supposons que pour tout $w \in \Sigma^*, |w| < l$, on ait que $w \in X$ si et seulement si $w \in K^*L$.

Soit $w \in \Sigma^*, |w| = l$.

Supposons que $w \in K^*L$. Alors, soit $w \in L$, et donc $w \in X$, soit $w \in K^+L$. Dans le dernier cas, il existe nécessairement $u \in K$ et $v \in K^*L$ tels que $w = uv$. Puisque $|u| > 0$, on a que $|v| < l$, et donc par hypothèse de récurrence, que $v \in X$. Ainsi donc, $w = uv \in KX \subseteq X$.

Inversement, supposons que $w \in X$. Alors, soit $w \in L$, et donc $w \in K^*L$, soit $w \in KX$. Dans le dernier cas, il existe nécessairement $u \in K$ et $v \in X$ tels que $w = uv$. Puisque $|u| > 0$, on a que $|v| < l$, et donc par hypothèse de récurrence, que $v \in K^*L$. Ainsi donc, $w = uv \in KK^*L \subseteq K^*L$.

Exercice 9.

Qu'advient-il du lemme précédent si K contient le mot vide ?

Solution 9.

Alors sont solutions tous les langages de la forme K^*M avec $L \subseteq M$.

Soit $X \subseteq \Sigma^*$ tel que $X = KX \cup L$. Soit $M = L \cup X \setminus (K \setminus \{\varepsilon\})^+X$. On peut montrer que pour tout $w \in \Sigma^*$, $w \in X$ si et seulement si $w \in K^*M$, par récurrence sur la longueur de w. D'où, nécessairement $X = K^*M$ pour un $M \subseteq \Sigma^*$ vérifiant $L \subseteq M$. En réalité, tout tel M convient, donc $X \in \{K^*M \mid M \subseteq \Sigma^*, L \subseteq M\}$.

Exercice 10.

Étant donné $n \in \mathbb{N}_{>0}$, soit $K_{i,j}$ pour $i, j \in [\![1, n]\!]$ ainsi que L_i pour $i \in [\![1, n]\!]$ des langages sur un alphabet Σ. Montrer que si, pour tous $i, j \in [\![1, n]\!]$, $\varepsilon \notin K_{i,j}$, alors il existe $X_1, X_2, \ldots, X_n \subseteq \Sigma^*$ uniques tels que

$$\begin{aligned}
X_1 &= K_{1,1}X_1 \cup K_{1,2}X_2 \cup \cdots \cup K_{1,n}X_n \cup L_1\\
X_2 &= K_{2,1}X_1 \cup K_{2,2}X_2 \cup \cdots \cup K_{2,n}X_n \cup L_2\\
\vdots & \qquad\qquad\qquad \vdots\\
X_n &= K_{n,1}X_1 \cup K_{n,2}X_2 \cup \cdots \cup K_{n,n}X_n \cup L_n \ ,
\end{aligned}$$

qui de plus sont rationnels dès que $K_{i,j}$ pour $i, j \in [\![1, n]\!]$ et L_i pour $i \in [\![1, n]\!]$ le sont.

Solution 10.

Montrons ce résultat par récurrence sur n.

Case de base $n = 1$. Il s'agit du lemme d'Arden.

Récurrence. Soit $n \in \mathbb{N}, n \geq 2$ et supposons que le résultat est vrai pour $n - 1$.

Soit $K_{i,j}$, avec $\varepsilon \notin K_{i,j}$, pour $i, j \in [\![1, n]\!]$ ainsi que L_i pour $i \in [\![1, n]\!]$ des langages sur un alphabet Σ. Nous allons montrer qu'il existe $X_1, X_2, \ldots, X_n \subseteq \Sigma^*$ uniques tels que

$$\begin{aligned}
X_1 &= K_{1,1}X_1 \cup K_{1,2}X_2 \cup \cdots \cup K_{1,n}X_n \cup L_1\\
X_2 &= K_{2,1}X_1 \cup K_{2,2}X_2 \cup \cdots \cup K_{2,n}X_n \cup L_2\\
\vdots & \qquad\qquad\qquad \vdots\\
X_n &= K_{n,1}X_1 \cup K_{n,2}X_2 \cup \cdots \cup K_{n,n}X_n \cup L_n
\end{aligned} \tag{1}$$

et qui de plus sont rationnels dès que $K_{i,j}$ pour $i, j \in [\![1, n]\!]$ et L_i pour $i \in [\![1, n]\!]$ le sont.

Soit $X_1, \ldots, X_{n-1} \subseteq \Sigma^*$ fixés. Puisque

$$K_{n,1}X_1 \cup K_{n,2}X_2 \cup \cdots \cup K_{n,n}X_n \cup L_n = K_{n,n}X_n \cup (K_{n,1}X_1 \cup \cdots \cup K_{n,n-1}X_{n-1} \cup L_n) \ ,$$

par le lemme d'Arden, il existe donc un unique $X_n \subseteq \Sigma^*$ tel que

$$X_n = K_{n,1}X_1 \cup K_{n,2}X_2 \cup \cdots \cup K_{n,n}X_n \cup L_n \ ,$$

qui est

$$X_n = K_{n,n}^*(K_{n,1}X_1 \cup \cdots \cup K_{n,n-1}X_{n-1} \cup L_n) \ .$$

Par conséquent, pour que $X_1, X_2, \ldots, X_n$ vérifient (1), il faut et il suffit que $X_1, \ldots, X_{n-1}$ vérifient

$$\begin{aligned} X_1 &= K'_{1,1}X_1 \cup \cdots \cup K'_{1,n-1}X_{n-1} \cup L'_1 \\ &\vdots \qquad\qquad \vdots \\ X_{n-1} &= K'_{n-1,1}X_1 \cup \cdots \cup K'_{n-1,n-1}X_{n-1} \cup L'_{n-1} \end{aligned} \tag{2}$$

où $K'_{i,j} = K_{i,j} \cup K_{i,n}K^*_{n,n}K_{n,j}$, qui ne peut pas contenir ε, pour $i, j \in [\![1, n-1]\!]$ et $L'_i = L_i \cup K_{i,n}K^*_{n,n}L_n$ pour $i \in [\![1, n-1]\!]$. Par hypothèse de récurrence, il existe ainsi $X_1, \ldots, X_{n-1} \subseteq \Sigma^*$ uniques vérifiant (2), qui de plus sont rationnels dès que $K'_{i,j}$ pour $i, j \in [\![1, n-1]\!]$ et L'_i pour $i \in [\![1, n-1]\!]$ le sont.

On peut en conclure que $X_1, X_2, \ldots, X_n \subseteq \Sigma^*$ vérifiant (1) sont uniques et sont de plus rationnels dès que $K_{i,j}$ pour $i, j \in [\![1, n]\!]$ et L_i pour $i \in [\![1, n]\!]$ le sont, puisqu'alors $K'_{i,j}$ pour $i, j \in [\![1, n-1]\!]$ et L'_i pour $i \in [\![1, n-1]\!]$ le sont aussi.

Exercice 11.

En déduire un algorithme permettant, étant donné un AFN $\mathcal{A}$, de construire une expression rationnelle correspondant à $\mathcal{L}(\mathcal{A})$.

Solution 11.

Soit $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$ un AFN.

Notons $Q = \{q_1, q_2, \ldots, q_n\}$ où $|Q| = n$. Posons $K_{i,j} = \{a \in \Sigma \mid q_j \in \delta(q_i, a)\}$ pour $i, j \in [\![1, n]\!]$ et $L_i = \begin{cases} \{\varepsilon\} & \text{si } q_i \in F \\ \emptyset & \text{sinon} \end{cases}$ pour $i \in [\![1, n]\!]$.

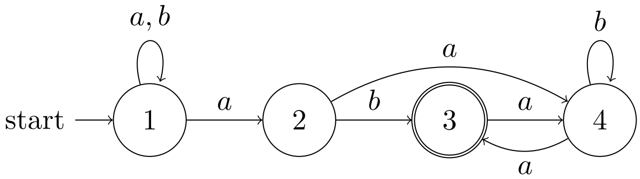
Posons à présent $R_i = \{w \in \Sigma^* \mid \hat{\delta}(q_i, w) \cap F \neq \emptyset\}$ pour tout $i \in [\![1, n]\!]$. On peut montrer que $R_1, R_2, \ldots, R_n$ vérifient

$$\begin{aligned} X_1 &= K_{1,1}X_1 \cup K_{1,2}X_2 \cup \cdots \cup K_{1,n}X_n \cup L_1 \\ X_2 &= K_{2,1}X_1 \cup K_{2,2}X_2 \cup \cdots \cup K_{2,n}X_n \cup L_2 \\ &\vdots \qquad\qquad\qquad \vdots \\ X_n &= K_{n,1}X_1 \cup K_{n,2}X_2 \cup \cdots \cup K_{n,n}X_n \cup L_n\ . \end{aligned} \tag{3}$$

Par le résultat précédent, $R_1, R_2, \ldots, R_n$ sont les uniques $X_1, X_2, \ldots, X_n$ vérifiant (3). Pour calculer algorithmiquement une expression rationnelle pour chacun d'eux, il suffit donc d'utiliser récursivement le procédé d'élimination basé sur le lemme d'Arden présenté dans la solution de l'exercice précédent, en observant qu'on peut très facilement écrire une expression rationnelle pour les $K_{i,j}$ et les L_i, qui sont ensuite combinés par addition, concaténation et étoile dans cette procédure. On en déduit donc une expression rationnelle correspondant à $\mathcal{L}(\mathcal{A})$.

Exercice 12.

Donner une expression rationnelle correspondant au langage reconnu par l'AFN suivant, sur $\{a, b\}$:



.

Solution 12.

On cherche $X_1, X_2, X_3, X_4 \subseteq \{a,b\}^*$ uniques tels que :

$$\begin{aligned} X_1 &= \{a,b\}^* X_1 \cup \{a\} X_2 \\ X_2 &= \{b\} X_3 \cup \{a\} X_4 \\ X_3 &= \{a\} X_4 \cup \{\varepsilon\} \\ X_4 &= \{a\} X_3 \cup \{b\} X_4 \ . \end{aligned}$$

On trouve alors, après élimination successive des lignes, que

$$\begin{aligned} X_1 &= \mathcal{L}((a+b)^* a(b+ab^*a)(ab^*a)^*) \\ X_2 &= \mathcal{L}((b+ab^*a)(ab^*a)^*) \\ X_3 &= \mathcal{L}((ab^*a)^*) \\ X_4 &= \mathcal{L}(b^*a(ab^*a)^*) \ . \end{aligned}$$

L'expression rationnelle $(a+b)^* a(b+ab^*a)(ab^*a)^*$ correspond donc au langage reconnu par l'AFN en question.

4 Caractérisations

Soit Σ un alphabet.

Exercice 13.

Donner une condition nécessaire et suffisante sur la forme d'une expression rationnelle E sur Σ pour que $\mathcal{L}(E)$ soit reconnu par un AFD $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ ne contenant aucun circuit non trivial (c'est-à-dire, tel qu'il n'existe pas $u, v \in \Sigma^*$ et $q, p \in Q, p \neq q$ vérifiant $\hat{\delta}(q,u) = p$ et $\hat{\delta}(p,v) = q$).

Solution 13.

Il s'agit de toutes les expressions rationnelles sommes d'expressions rationnelles de la forme

$$A_0^* a_1 A_1^* a_2 \cdots A_{k-1}^* a_k A_k^*$$

où $k \in \mathbb{N}$, $A_0, A_1, \ldots, A_k \subseteq \Sigma$ et $a_1, a_2, \ldots, a_k \in \Sigma$ tels que $a_i \notin A_{i-1}$ pour tout $i \in [\![1, k]\!]$.

Exercice 14.

Donner une condition nécessaire et suffisante sur la forme d'une expression rationnelle E sur Σ pour que $\mathcal{L}(E)$ soit reconnu par un AFD $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ tel que pour tous $u, v \in \Sigma^*$ et $q \in Q$, $\hat{\delta}(q, uv) = \hat{\delta}(q, vu)$.

Solution 14.

Soit $\Sigma = \{a_1, \ldots, a_s\}$ pour $s \in \mathbb{N}_{>0}$.

Étant donnés $k \in \mathbb{N}$ et $p \in \mathbb{N}_{>0}$, on définit la relation d'équivalence $\sim_{k,p}$ sur $\mathbb{N}$ telle que pour tous $i, j \in \mathbb{N}$, $i \sim_{k,p} j$ si et seulement si :

— $i < k$ et $i = j$; ou

— $i \geq k$ et $j \geq k$ et $i \equiv j \mod p$.

Pour tous $i_1, \ldots, i_s \in \mathbb{N}$, on définit alors le langage

$$F_{k,p}(i_1, \ldots, i_s) = \{w \in \Sigma^* \mid \forall j \in [\![1, n]\!], |w|_{a_j} \sim_{k,p} i_j\} \ ,$$

qui est rationnel puisque pour tout $j \in [\![1, s]\!]$, $\{w \in \Sigma^* \mid |w|_{a_j} \sim_{k,p} i_j\}$ est rationnel. Pour tous $i_1, \ldots, i_s \in \mathbb{N}$, il existe alors une expression rationnelle $E_{k,p}(i_1, \ldots, i_s)$ sur Σ telle que $\mathcal{L}(E_{k,p}(i_1, \ldots, i_s))$ $= F_{k,p}(i_1, \ldots, i_s)$.

Il s'agit donc de toutes les expressions rationnelles sommes, pour $k \in \mathbb{N}$ et $p \in \mathbb{N}_{>0}$ fixés, d'expressions rationnelles $E_{k,p}(i_1, \ldots, i_s)$ pour $i_1, \ldots, i_s \in \mathbb{N}$.

Exercice 15.

Donner une condition nécessaire et suffisante pour qu'un AFD $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ vérifie que $\mathcal{L}(\mathcal{A})$ est une combinaison booléenne de langages de la forme $\Sigma^* a \Sigma^*$ avec $a \in \Sigma$ (c'est-à-dire, une union d'intersections de langages de cette forme ou leurs compléments).

Solution 15.

Il s'agit de tous les AFD $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ tels que pour tous $u, v \in \Sigma^*$ et $q \in Q$, on ait d'une part $\hat{\delta}(q, u^2) = \hat{\delta}(q, u)$, et d'autre part $\hat{\delta}(q, uv) = \hat{\delta}(q, vu)$.