le cnam

Codes et Automates finis

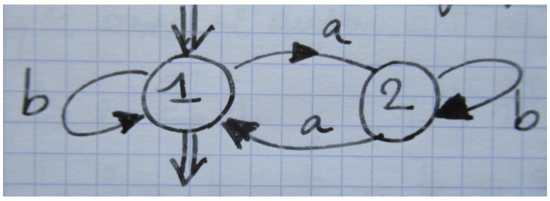
Cours 10 Automates finis déterministes

• Définition

Un automate fini déterministe est la donnée du quintuplet $\mathscr{A} = (Q, A, T, I, F)$ où l'ensemble des états Q est un ensemble fini, l'alphabet A est un ensemble fini de lettres, la fonction de transition de l'automate est notée T, l'ensemble des états initiaux est désigné par I, avec $I \subset Q$ et $I \neq \emptyset$, et l'ensemble des états finals est nommé F, avec $F \subset Q$ et $F \neq \emptyset$.
La transition d'un automate déterministe est définie à l'aide d'une application
$\delta : Q \times A \longrightarrow Q$, qui à tout couple $(p, a) \in Q \times A$, associe un unique état $q = \delta(p, a)$. La fonction de transition est alors l'ensemble des triplets T de la forme $(p, a, \delta(p, a))$ pour $p \in Q$ et $a \in A$: $T = \{(p, a, \delta(p, a)), p \in Q, a \in A\}$.
Un exemple d'automate fini : $Q = \{1, 2\}$, $A = \{a, b\}$,
$T = \{(1, a, 2), (1, b, 1), (2, a, 1), (2, b, 2)\}$, $I = \{1\}$, $F = \{1\}$.

• Représentation graphique d'un automate

On note les états $p \in Q$ sous la forme $ⓟ$; une transition $q = \delta(p, a)$ est représentée avec une flèche $ⓟ \xrightarrow{a} ⓠ$ qui exprime qu'à la lecture de la lettre a, l'état actif de l'automate se transporte de l'état p à l'état q. Un état initial est représenté avec une double flèche entrante : $\Longrightarrow ⓘ$. Un état final est représenté avec une double flèche sortante : $ⓕ \Longrightarrow$. L'exemple précédent se résume à l'aide du graphe ci-dessous.



Dans une transition $q = \delta(p, a)$, la lettre a est une étiquette, un label ou un ordre de transfert pour passer de l'état p à l'état q.

• Chemin dans un automate

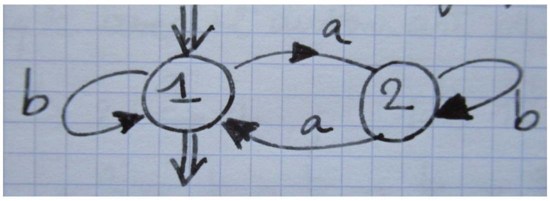
Un chemin $q_0, q_1, \ldots, q_n$ dans l'automate $\mathscr{A}$ est une suite finie de transitions :
$q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \xrightarrow{a_3} \ldots q_{n-1} \xrightarrow{a_n} q_n$, avec $q_j \in Q$ et $a_j \in A$. L'état $q_0 \in Q$ est l'état de départ du chemin et $q_n \in Q$ l'état d'arrivée. On peut aussi noter ce chemin $q_0 \xrightarrow{a_1 a_2 \ldots a_n} q_n$ et $m = a_1 a_2 \ldots a_n$ est un mot formé avec des lettre de l'alphabet A. Le mot $m \in A^*$ pilote le chemin de l'état q_0 à l'état q_n en suivant à chaque étape une transition permise : $q_j = \delta(q_{j-1}, a_j)$ pour $1 \leq j \leq n$.

François Dubois, avril 2020.

Avec l'automate décrit à la figure ci-dessus, les chemins $1 \xrightarrow{a} 2 \xrightarrow{a} 1$, $2 \xrightarrow{a} 1 \xrightarrow{b} 1 \xrightarrow{a} 2$, $1 \xrightarrow{b} 1 \xrightarrow{b} 1 \xrightarrow{a} 2 \xrightarrow{b} 2$ ou $1 \xrightarrow{a} 2 \xrightarrow{b} 2 \xrightarrow{b} 2 \xrightarrow{b} 2 \xrightarrow{a} 1 \xrightarrow{b} 1$ sont des chemins compatibles avec les transitions de l'automate.

- Chemin acceptant

Un chemin acceptant dans l'automate $\mathscr{A} = (Q, A, T, I, F)$ est un chemin $q_0 \xrightarrow{a_1 a_2 \dots a_n} q_n$ tel que $n \in \mathbb{N}$ est un entier naturel, q_0 est un état initial : $q_0 \in I$ et q_n est un état final : $q_n \in F$.
Ainsi, avec l'automate décrit page 1 et reproduit ci-dessous, les chemins $1 \xrightarrow{b} 1$, $1 \xrightarrow{a} 2 \xrightarrow{a} 1$, $1 \xrightarrow{b} 1 \xrightarrow{a} 2 \xrightarrow{a} 1 \xrightarrow{b} 1$ sont des chemins acceptants : on part de l'état initial ("1" dans cet exemple) et on arrive à l'état final (encore égal à "1" pour l'automate de la page 1).



- Mot accepté par l'automate

Un mot $m \in A^*$ est accepté par l'automate $\mathscr{A} = (Q, A, T, I, F)$ si m pilote un chemin acceptant : $q_0 \in I$, $q_n \in F$, $q_0 \xrightarrow{m} q_n$. L'entier $n = |m|$ est le nombre de lettres du mot m, c'est à dire la longueur de m.
Par exemple pour l'automate de la figure ci-dessus et les chemins acceptants proposés au paragraphe précédent, les mots $m = b$, $m = aa$ et $m = baab$ sont des mots acceptés par l'automate.

- Extension au mot sans lettre

Dans le cas de la figure ci-dessus où l'état initial est aussi l'état final, on a envie d'étendre la notion de transition au mot sans lettre : $1 \xrightarrow{\varepsilon} 1$, ce qui permet de considérer le mot sans lettre ε comme un mot accepté par l'automate.
De façon générale, le mot sans lettre pilote un chemin de longueur nulle $p \xrightarrow{\varepsilon} p$ qui consiste à ne pas se déplacer dans l'automate.

- Langage $\mathscr{L}(\mathscr{A})$ d'un automate fini $\mathscr{A}$

Étant donné un automate fini $\mathscr{A} = (Q, A, T, I, F)$, l'ensemble de tous les mots acceptés par l'automate constitue un langage sur l'alphabet A, noté $\mathscr{L}(\mathscr{A})$; c'est le langage de l'automate fini $\mathscr{A}$.
Par exemple, pour l'automate fini de la figure ci-dessus, nous allons démontrer que
$\mathscr{L}(\mathscr{A}) = (b + ab^* a)^*$.

- Langage rationnel

Un langage rationnel sur un alphabet A est un langage L qui peut être l'un des langages très simples suivants : $L = \emptyset$ ou $L = \{a\}$ (une lettre de l'alphabet). Il peut aussi être obtenu par toute combinaison des trois opérations fondamentales suivantes : union (ou addition), concaténation (ou produit), étoile. Rappelons que $\emptyset^* = \varepsilon$ et que pour un langage L,
$L^* = \{\varepsilon\} + L + L^2 + L^3 + \dots$.

Si L et L' sont deux langages rationnels, la réunion $L+L'=L\cup L'$ est aussi un langage rationnel, le produit de concaténation $L.L'=\{mm', m\in L, m'\in L'\}$ est encore un langage rationnel. Enfin, si L est un langage rationnel, l'étoile L^* est aussi un langage rationnel.
Ainsi, le langage $L=(b+ab^*a)^*$ est un langage rationnel. En effet, $\{b\}$ est rationnel, donc b^* l'est aussi. De même, $\{a\}$ est rationnel, donc ab^* puis ab^*a sont des langage rationnels. La réunion $b+ab^*a$ est donc un langage rationnel et quand on considère son étoile de Kleene, on obtient un nouveau langage rationnel.

- Théorème de Kleene (1956)

Un langage L est rationnel si et seulement si il existe un automate fini déterministe $\mathscr{A}$ tel que $\mathscr{L}(\mathscr{A})=L$.
La preuve de ce résultat est longue et délicate ; elle ne sera pas donnée dans le cadre de ce cours. Nous renvoyons le lecteur curieux au livre d'Olivier Carton : *Langages formels ; calculabilité et complexité*, Vuibert, 2014. Nous retenons surtout les deux questions pratiques induites par le théorème de Kleene :
Question 1. Pour un automate fini déterministe donné $\mathscr{A}$, comment déterminer, calculer, le langage rationnel $\mathscr{L}(\mathscr{A})$ de ses mots acceptés ? La réponse est développée dans la suite de cette leçon.
Question 2. Si on se donne un langage rationnel L, déterminer un (nous verrons qu'il peut y en avoir plusieurs) automate fini déterministe $\mathscr{A}=(Q,A,T,I,F)$ de sorte que $\mathscr{L}(\mathscr{A})=L$. Seul l'alphabet A est connu puisque le langage L est donné ; la réponse à cette seconde question occupera plusieurs leçons dans la suite de ce cours.

- Langage associé à un état de l'automate

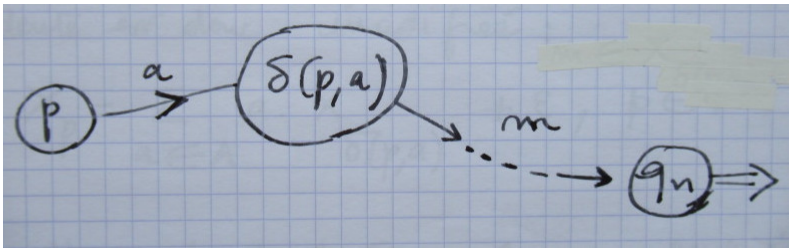
Afin de calculer le langage $\mathscr{L}(\mathscr{A})$ de l'automate fini $\mathscr{A}$, on introduit toute une famille de langages X_p paramétrée par tous les états $p\in Q$ de l'automate. On définit X_p comme l'ensemble des mots qui étiquettent un chemin de l'automate $\mathscr{A}$ permettant d'aller de l'état p à un état final $q_n\in F$: $X_p=\{m\in A^*,\ p=q_0\xrightarrow{m}q_n,\ q_n\in F\}$.
Avec la figure de la page précédente, on a par exemple $X_2=\{a, ba, bba, aaa, ab, \dots\}$,
$X_1=\{aa, b, bb, baba, \dots\}$. On remarque que la particularité $I=F=\{1\}$ de ce cas permet d'écrire que le mot sans lettre appartient au langage X_1.

- Langage de l'automate $\mathscr{A}$

Avec les notations précédentes, $\mathscr{L}(\mathscr{A})=\cup_{i\in I}X_i$.
En effet, un mot accepté par l'automate $\mathscr{A}$ pilote un chemin qui part d'un étal initial $i\in I$ et se termine par un état final $q_n\in F$.

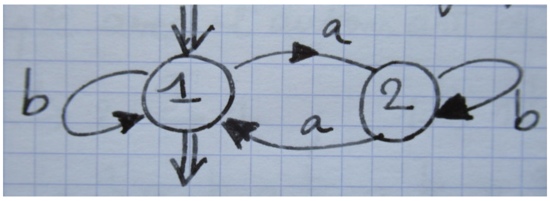
- Équations de départ

Nous allons établir tout un ensemble d'équations entre les langages X_q pour $q \in Q$, avec les opérations de réunion, concaténation et l'étoile. Le raisonnement (voir la figure ci-dessous) consiste à dire que si un chemin part de l'état $p \in Q$, il doit tout d'abord atteindre l'un des états $q = \delta(p, a)$ pour une lettre $a \in A$. Puis de cet état fixé, il suffit de trouver un chemin qui va de $\delta(p, a)$ à un état final comme sur la figure ci-dessous. Une étiquette qui pilote un tel chemin est donc de la forme $a.m$, où $a \in A$ est une lettre de l'alphabet et m une étiquette qui pilote un chemin allant de $\delta(p, a)$ à $q_n \in F$, c'est à dire $m \in X_{\delta(p,a)}$.



Si $p \notin F$, une étiquette qui pilote un chemin de p à un état final appartient au langage $a X_{\delta(p,a)}$ pour une certaine lettre a de l'alphabet A. Donc en faisant la réunion de tous les cas possibles, on a la relation $X_p \subset \cup_{a\in A} \big(a X_{\delta(p,a)}\big)$. Réciproquement, si un mot m' s'écrit sous la forme $m' = am$, avec $m \in X_{\delta(p,a)}$, alors il étiquette un chemin qui va de l'état p à un état final. On peut écrire $X_p = \sum_{a\in A} \big(a X_{\delta(p,a)}\big)$ si $p \notin F$.

Si $p \in F$, on est déjà sur un état final. Le mot sans lettre ε permet alors également d'aller à un état final (!). La relation précédente doit donc être modifiée afin de prendre en compte cette remarque : $X_p = \sum_{a\in A} \big(a X_{\delta(p,a)}\big) + \varepsilon$ si $p \in F$.



- Étude d'un exemple

Pour l'exemple considéré depuis le début de ce chapitre (décrit à la figure ci-dessus), on peut écrire ces deux équations qui relient X_1 et X_2 ; $X_1 = aX_2 + bX_1 + \varepsilon$, $X_2 = bX_2 + aX_1$.

On résout le système linéaire dit des "équations à gauche" en utilisant le lemme d'Arden. Pour l'exemple de la figure ci-dessus, on extrait X_2 de la seconde équation, de la forme $X = K.X + L$ avec $K = \{b\} \not\ni \varepsilon$ c'est à dire ici $X_2 = b^* a X_1$. On reporte cette valeur du langage X_2 dans la première équation : $X_1 = a(b^* a X_1) + bX_1 + \varepsilon = (ab^*a + b) X_1 + \varepsilon$. Cette équation est encore de la forme $X = K.X + L$ avec $K = ab^*a + b$ cette fois. Le langage $ab^*a + b$ ne contient pas le mot sans lettre. Donc $X = K^* L$ qui s'écrit ici $X_1 = (ab^*a + b)^* \varepsilon = (ab^*a + b)^*$.

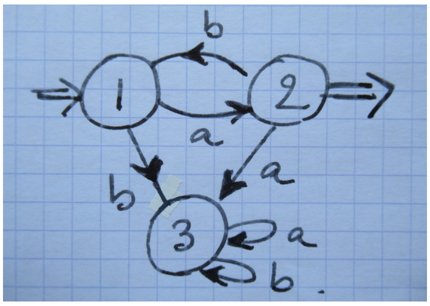
Comme l'état "1" est aussi l'unique état initial de l'automate, on a

$\mathscr{L}(\mathscr{A}) = \cup_{i\in I} X_i = X_1 = (ab^*a + b)^*$, ainsi qu'annoncé plus haut.

Exercices

- Étude d'un automate fini déterministe

On étudie l'automate fini décrit à la figure ci-dessous.



a) Les mots *ab*, *aa*, *ba*, *a*, *aba*, *aaba* sont-ils reconnus par l'automate ?

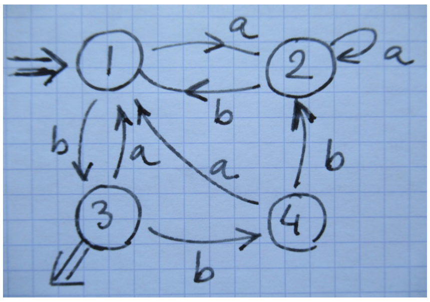
b) Ecrire les équations de départ liant les trois langages X_1, X_2 et X_3.

c) Montrer que $X_3 = \emptyset$.

d) En déduire une expression du langage de cet automate.

- Un autre automate fini déterministe

On se donne un automate décrit à la figure ci-dessous.



a) Les mots *b*, *ab*, *bbbb* *aaaa*, *bbab* sont-ils reconnus par l'automate ?

b) Écrire les équations de départ reliant les langages X_1, X_2, X_3 et X_4.

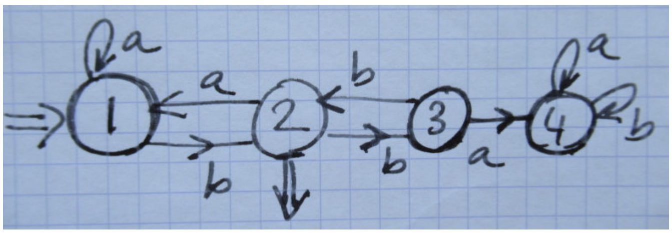
c) Exprimer X_2 en fonction de X_1 à l'aide du lemme d'Arden.

d) Achever la résolution de ce système d'équations.

e) Quel est le langage de cet automate ?

- Un automate fini déterministe

Dans cet exercice, on s'intéresse à l'automate fini représenté ci-dessous.



a) Quels sont les mots de longueur inférieure ou égale à 3 reconnus par cet automate ?

b) Le mot $m = bbbaabbaabbbaabbaabbb$ est-il reconnu par l'automate ?

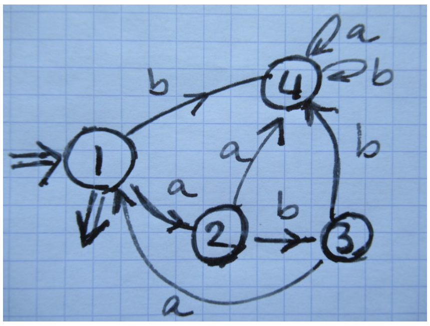
c) Écrire les équations de départ reliant les langages X_1, X_2, X_3 et X_4.

d) Résoudre ce système d'équations.

e) Quel est le langage de cet automate ?

• Un quatrième automate fini déterministe

Dans cet exercice, on s'intéresse à l'automate fini représenté ci-dessous.



a) Les mots *ababab*, *bababa*, *abaabaaba* et *aabbaa* sont-ils reconnus par cet automate ?

b) Écrire le système des équations de départ.

c) Résoudre ce système d'équations.

d) Quel est le langage $\mathscr{L}(\mathscr{A})$ de cet automate ?