R4.A.12 Automates et Langages

Thibault Godin ; Lucie Naert

IUT Vannes, Département informatique

Motivation

Comment vérifier *effectivement* que $\boldsymbol{u} = abbba$ appartient au langage $\mathcal{L} = ab^*a$ mais que $\boldsymbol{v} = abbab$ n'y appartient pas ?

DFA

Un **automate déterministe fini** est un quintuplet $A = (\Sigma, Q, i_0, F, \delta)$, où :

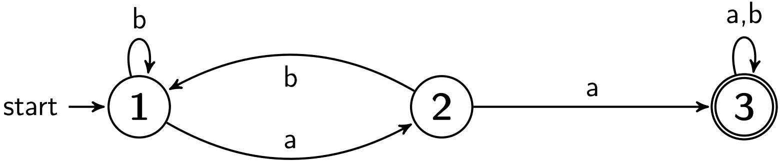
- Σ est un ensemble fini, appelé **alphabet**.
- Q est un ensemble fini, appelé **ensemble des états**. (Q et Σ sont disjoints)
- i_0 est un élément distingué de Q, appelé **état initial**.
- F est un sous-ensemble de Q, appelé **ensemble des états finaux**.
- δ est une application de $Q \times \Sigma$ dans Q, appelée **fonction de transition**.

Exemple : Considérons l'automate $A = (\Sigma, Q, i_0, F, \delta)$ suivant : $\Sigma = \{a, b\}$, $Q = \{1, 2, 3\}$, $i_0 = 1$, $F = \{3\}$, $\delta(1, a) = 2$, $\delta(1, b) = 1$, $\delta(2, a) = 3$, $\delta(2, b) = 1$, $\delta(3, a) = 3$ et $\delta(3, b) = 3$.

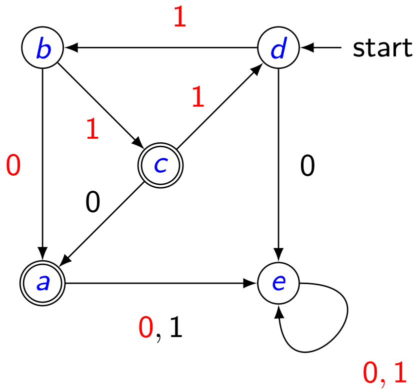
On peut écrire la **table de transition** de δ :

δ	a	b
1	2	1
2	3	1
3	3	3

On dessine le **graphe de transition** de l'automate :



Lecture et acceptation de mots



$$A = (\Sigma, Q, i_0, F, \delta)$$

$$d \xrightarrow{1} b \xrightarrow{0} a \xrightarrow{0} e \xrightarrow{0} e \xrightarrow{1} e$$

$$d \xrightarrow{0} b \xrightarrow{1} c \xrightarrow{1} d \xrightarrow{1} b \xrightarrow{0} a$$

$\boldsymbol{u} \in \mathcal{L}_A$ ssi le mot $\boldsymbol{\delta}_u \in Q^*$ de longueur[1] $|\boldsymbol{u}| + 1$ défini par $\boldsymbol{\delta}_u[0] = i_0$ et $\boldsymbol{\delta}_u[k+1] = \delta(\boldsymbol{\delta}_u[k], \boldsymbol{u}[k+1])$ est tel que $\boldsymbol{\delta}_u[|\boldsymbol{u}|] \in F$

$\mathcal{L}_A$ est le langage reconnu par l'automate A.

1. On fait ici démarrer l'indexation de $\boldsymbol{u}$ à 1 et celle de $\boldsymbol{\delta}_u$ à 0 pour améliorer la lisibilité

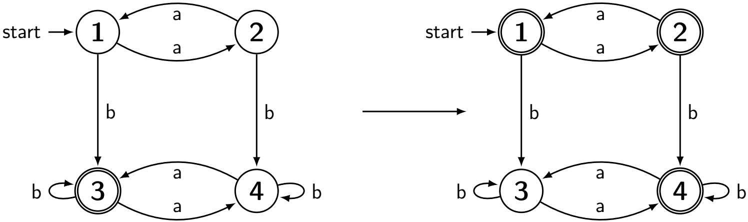
Langage reconnaissables déterministes

$\rightsquigarrow$ classe de langages reconnus par un DFA : les langages *reconnaissables par automates déterministes*, *i.e.* $L \in \mathcal{L}_{det\ rec} \iff \exists A$ DFA , $L = L_A$

Montrer que la classe $\mathcal{L}_{det\ rec}$ est close pour l'intersection,le complément et le préfixe.
En déduire qu'elle est close pour l'union.

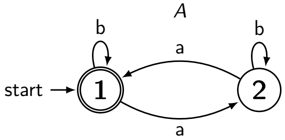
	$\mathcal{L}_{rec\ det}$
Union	clos
Intersection	clos
Concaténation	?
Complément	clos
Préfixe	clos
Suffixe	?
Miroir	?
Étoile	?

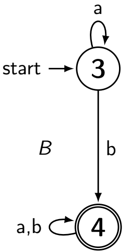
Construction : complément

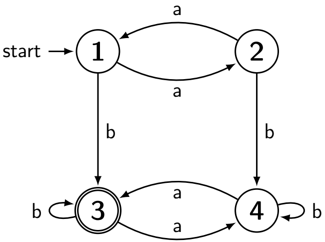


Si $A = (\Sigma, Q, i_0, F, \delta)$ accepte le langage L alors $A_c = (\Sigma, Q, i_0, Q \setminus F, \delta)$ accepte $\bar{L} = \Sigma^* \setminus L$

Construction : intersection

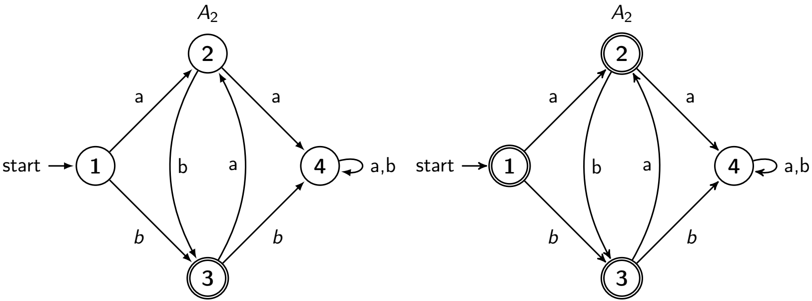






Si $A = (\Sigma, Q, i_0, F, \delta)$ accepte le langage L et $B = (\Sigma, Q', j_0, F', \delta')$ accepte le langage M alors $AB = (\Sigma, Q \times Q', (i_0, j_0), Q \times Q', \Delta)$ accepte $L \cap M$ (avec $\Delta((i,j), a) = (\delta(i,a), \delta'(j,a))$ la fonction de transition produit)

Construction : préfixe



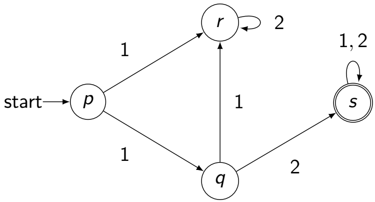
Un état q devient acceptant s'il est *co-accessible* depuis un état final, c-à-d si on peut atteindre un état final depuis q.

NFA

Un **automate non-déterministe fini** est un quintuplet $A = (\Sigma, Q, \delta, I, F)$, où :

- Σ est un ensemble fini, appelé **alphabet**.
- Q est un ensemble fini, appelé **ensemble des états**. (Q et Σ sont disjoints)
- I est un sous-ensemble de Q, appelé **ensemble des états initiaux**.
- F est un sous-ensemble de Q, appelé **ensemble des états finaux**.
- δ est une application de $Q \times \Sigma$ dans $\mathcal{P}(Q)$, appelée **fonction de transition**.

NFA



$A = (\{1,2\}, \{p,q,r,s\}, \delta, \{p\}, \{s\})$

	1	2
p	$\{r,q\}$	$\emptyset$
q	$\{r\}$	$\{s\}$
r	$\emptyset$	$\{r\}$
s	$\{s\}$	$\{s\}$

Un *run* associé à un mot $\boldsymbol{u}$ est le mot $\boldsymbol{\delta}_u \in Q^*$ de longueur [2] $|\boldsymbol{u}| + 1$ défini par $\boldsymbol{\delta}_u[0] \in I$ et $\boldsymbol{\delta}_u[k+1] \in \delta(\boldsymbol{\delta}_u[k], \boldsymbol{u}[k+1])$.

Un *run* (chemin) est *acceptant* si $\boldsymbol{\delta}_u[|\boldsymbol{u}|] \in F$

Un mot est accepté s'il existe un run acceptant associé à ce mot

$\mathcal{L}_A$ est le langage reconnu par l'automate A.

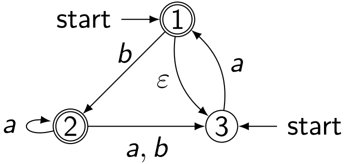
2. On fait ici démarrer l'indexation de $\boldsymbol{u}$ à 1 et celle de $\boldsymbol{\delta}_u$ à 0 pour améliorer la lisibilité

ε-NFA

Un **automate non-déterministe fini à transitions spontanées** est un quintuplet $A = (\Sigma, Q, \delta, I, F)$, où :

- Σ est un ensemble fini, appelé **alphabet**.
- Q est un ensemble fini, appelé **ensemble des états**. (Q et Σ sont disjoints)
- I est un sous-ensemble de Q, appelé **ensemble des états initiaux**.
- F est un sous-ensemble de Q, appelé **ensemble des états finaux**.
- δ est une application de $Q \times (\Sigma \cup \{\varepsilon\})$ dans $\mathcal{P}(Q)$, appelée **fonction de transition**.

ε-NFA



$$A = (\{a, b\}, \{1, 2, 3\}, \delta, \{1, 3\}, \{1, 2\})$$

	a	b	ε
1	$\emptyset$	$\{2\}$	$\{3\}$
2	$\{2, 3\}$	$\{3\}$	$\emptyset$
3	$\{1\}$	$\emptyset$	$\emptyset$

ε-Clôture

Soit $A = (\Sigma, Q, \delta, I, F)$ un ε-NFA. On peut obtenir un NFA équivalent $A' = (\Sigma, \hat{Q}, \delta^*, I, F)$

Pour cela on calcule les ε-clôtures, c-à-d l'ensemble des états que l'on peut atteindre par un nombre quelconque de ε-transitions.

La clôture $\hat{q}$ de l'état q est le plus petit ensemble décrit par $q \in \hat{q}$ et $\hat{q} = \bigcup_{p \in \delta(q,\varepsilon)} \hat{p}$

Attention, $\hat{q}$ est un ensemble ! $\hat{Q}$ sont les ε-clôtures δ^* donné par $\delta^*(\hat{q}, x) = \bigcup_{p \in \hat{q}} \delta(p, x)$

$A = (\{a, b\}, \{1, 2, 3\}, \delta, \{1, 3\}, \{1, 2\})$

	a	b	ε
1	$\emptyset$	$\{2\}$	$\{3\}$
2	$\{2, 3\}$	$\{3\}$	$\emptyset$
3	$\{1\}$	$\emptyset$	$\emptyset$

$\hat{1} = \{1, 3\}, \hat{2} = \{2\}, \hat{3} = \{3\}$

$A' = (\mathcal{P}(\{1, 2, 3\}), \{a, b\}, \delta, \{1, 3\}, \{1, 2\})$

	a	b
$\{1, 3\}$	$\{1, 3\}$	$\{2\}$
$\{2\}$	$\{2, 3\}$	$\{3\}$
$\{3\}$	$\{1, 3\}$	$\emptyset$

Langage reconnaissables non-déterministes

$\rightsquigarrow$ classe de langages reconnus par un (ε-)NFA : les langages *reconnaissables par automates non-déterministes*, *i.e.* $L \in \mathcal{L}_{ndet\ rec} \iff \exists A$ (ε-)NFA , $L = L_A$

Montrer que la classe $\mathcal{L}_{ndet\ rec}$ est close pour l'union, la concaténation et l'étoile.

	$\mathcal{L}_{ndet\ rec}$
Union	clos
Intersection	?
Concaténation	clos
Complément	?
Préfixe	?
Suffixe	?
Miroir	?
Étoile	clos

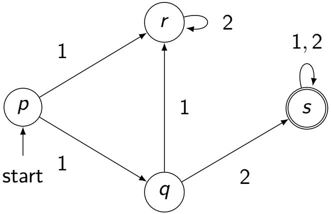
Déterminisation : $\mathcal{L}_{det\ rec} = \mathcal{L}_{ndet\ rec}$

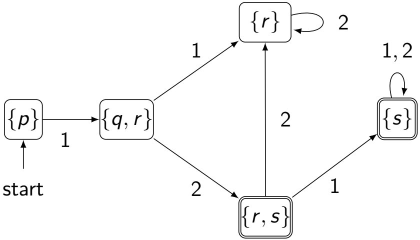
Rabin–Scott powerset construction :

À partir du NFA $N = (\Sigma, Q, \delta, I, F)$, on construit le DFA $D = (\Sigma, Q_d, \delta_d, i_d, F_d)$

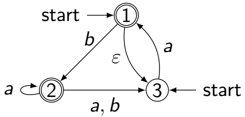
- $Q_d = \mathcal{P}(Q)$
- $i_d = I$
- $F_d = \{q \in Q_d | q \cap F \neq \emptyset\}$
- $\delta_d(q, x) = \bigcup_{p \in q} \delta(p, x) \quad q \in Q_d$

Déterminisation : $\mathcal{L}_{det\ rec} = \mathcal{L}_{ndet\ rec}$





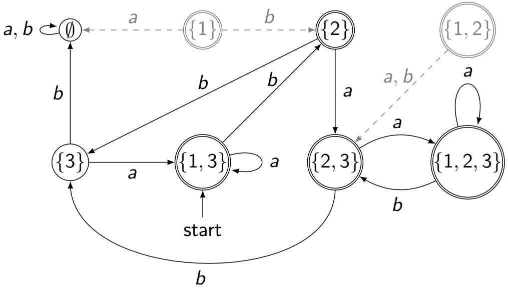
Déterminisation : $\mathcal{L}_{det\ rec} = \mathcal{L}_{ndet\ rec}$



	a	b	ε
1	$\emptyset$	$\{2\}$	$\{3\}$
2	$\{2,3\}$	$\{3\}$	$\emptyset$
3	$\{1\}$	$\emptyset$	$\emptyset$

	a	b
$\{1,3\}$	$\{1\}$	$\{2\}$
$\{2\}$	$\{2,3\}$	$\{3\}$
$\{3\}$	$\{1\}$	$\emptyset$

	a	b
$\{1,3\}$	$\{1,3\}$	$\{2\}$
$\{2\}$	$\{2,3\}$	$\{3\}$
$\{2,3\}$	$\{1,2,3\}$	$\{3\}$
$\{3\}$	$\{1,2\}$	$\emptyset$
$\{1,2,3\}$	$\{1,2,3\}$	$\{2,3\}$
$\emptyset$	$\emptyset$	$\emptyset$



Langages reconnaissables

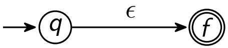
$\rightsquigarrow$ classe de langages reconnus par un automate finis : les langages *reconnaissables*, *i.e.* $L \in \mathcal{L}_{rec} \iff \exists A$ FA, $L = L_A$

	$\mathcal{L}_{ndet\ rec}$
Union	clos
Intersection	clos
Concaténation	clos
Complément	clos
Préfixe	clos
Suffixe	clos
Miroir	clos
Étoile	clos

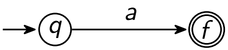
Ainsi $\mathcal{L}_{reg} \subset \mathcal{L}_{rec}$

$\mathcal{L}_{reg} \subset \mathcal{L}_{rec}$ Thompson (1968)[3]

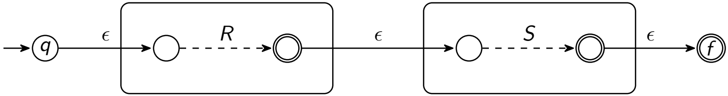
(A) Ensemble vide $\emptyset$



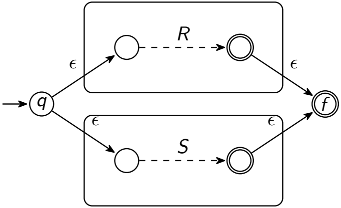
(B) Mot vide ϵ



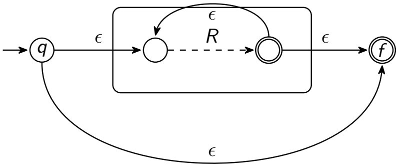
(C) Lettre $a \in \Sigma$



(D) Concaténation $R.S$



(E) Union $R \cup S$

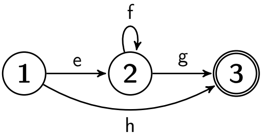


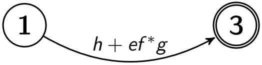
(F) Étoile R^*

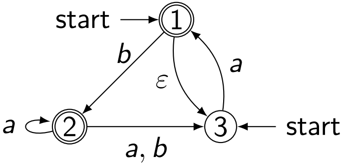
3. McNaugton–Yamada 1960

$\mathcal{L}_{reg} = \mathcal{L}_{rec}$

Brzozowski & McCluskey (1963)

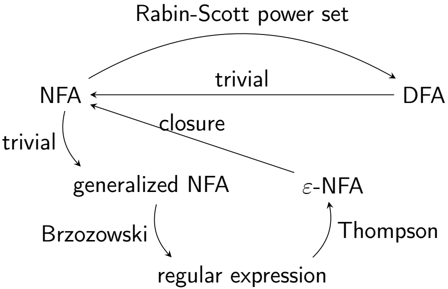






Ainsi $\mathcal{L}_{rec} \subset \mathcal{L}_{reg}$

Synthèse



	$\mathcal{L}_{rec}$
Union	clos
Intersection	clos
Concaténation	clos
Complément	clos
Préfixe	clos
Suffixe	clos
Miroir	clos
Étoile	clos