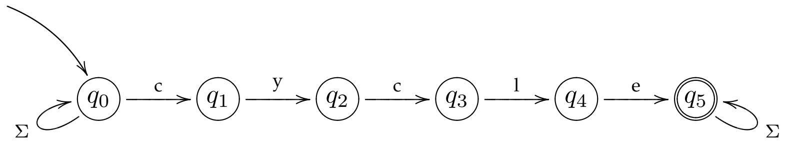
Informatique théorique III
Série 3 solution

Prof. Nestmann, 2004

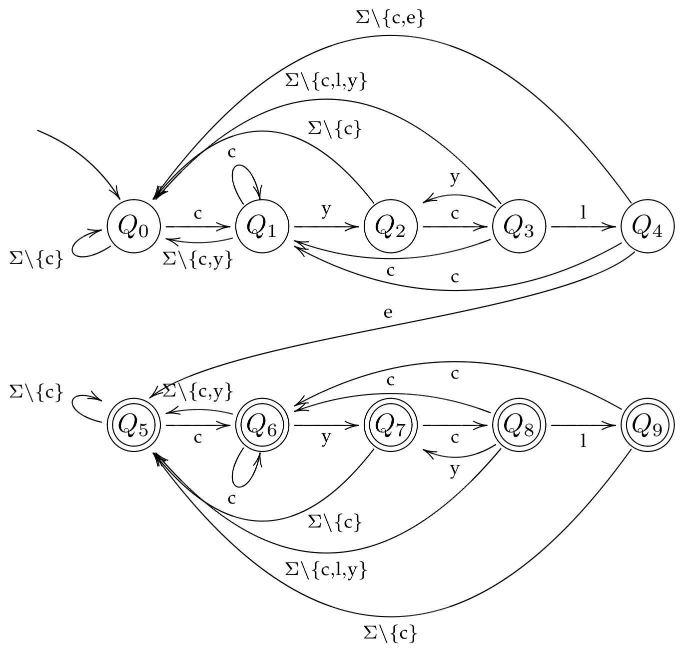
1. Recherche de mots-clef dans un texte

Soit Σ l'alphabet français. Un problème couramment rencontré en informatique est la recherche efficace de mots clefs dans un texte. Donnez le graphe d'un AFN M_v sur Σ pour aider un gérant de vélodrome à déterminer si un texte w contient au moins une occurence du mot « cycle ». Déterminiser ensuite l'AFN M_v et donner son graphe. □

Solution. On fournit au gérant l'AFN suivant :



On le comble de bonheur en lui donnant la version déterministe suivante :



avec $Q_0 \triangleq \{q_0\}$, $Q_1 \triangleq \{q_0, q_1\}$, $Q_2 \triangleq \{q_0, q_2\}$, $Q_3 \triangleq \{q_0, q_1, q_3\}$, $Q_4 \triangleq \{q_0, q_4\}$, $Q_5 \triangleq \{q_0, q_5\}$, $Q_6 \triangleq \{q_0, q_1, q_5\}$, $Q_7 \triangleq \{q_0, q_2, q_5\}$, $Q_8 \triangleq \{q_0, q_1, q_3, q_5\}$, $Q_9 \triangleq \{q_0, q_4, q_5\}$.

Notez que la présence de la "deuxième copie" constituée d'états finaux est dûe à l'application pure et simple de l'algorithme de déterminisation. On verra un peu plus tard un algorithme permettant de minimiser l'ensemble d'état d'un AFD. ■

2. Ordres produits

On considère dans cet exercice l'ensemble des entiers naturels $\mathbb{N}$.
On définit les relations binaires R_1 et R_2 sur $\mathbb{N} \times \mathbb{N}$ (les couples d'entiers) de la manière suivante :

$$(x, y)R_1(x', y') \Leftrightarrow x \leq x' \wedge y \leq y'$$
$$(x, y)R_2(x', y') \Leftrightarrow x < x' \vee (x = x' \wedge y \leq y')$$

1. Montrer que R_1 est une relation d'ordre.
2. Montrer que R_1 n'est pas une relation d'ordre totale.
3. Montrer que R_2 est une relation d'ordre.
4. Montrer que R_2 est une relation d'ordre totale.
5. Comparer R_1 et R_2 au sens de l'inclusion. Justifier la réponse. (est-ce que R_1 est un raffinement de R_2 ? est-ce que R_2 est un raffinement de R_1 ? ...)

$\square$

Solution.

1. Montrons que R_1 est une relation d'ordre :
   – réflexivité :
     Soit $(x, y) \in \mathbb{N} \times \mathbb{N}$. Montrons que $(x, y)R_1(x, y)$.
     Comme $\leq$ est réflexive, on a $x \leq x$ et $y \leq y$ donc $(x, y)R_1(x, y)$.
     Donc R_1 est réflexive.
   – transitivité :
     Soit $(x, y), (x', y'), (x'', y'') \in \mathbb{N} \times \mathbb{N}$.
     Supposons que $(x, y)R_1(x', y')$ et $(x', y')R_1(x'', y'')$. Montrons que $(x, y)R_1(x'', y'')$.
     Par définition de R_1, on a $x \leq x'$, $y \leq y'$, $x' \leq x''$ et $y' \leq y''$.
     Comme $\leq$ est transitive, on en déduit $x \leq x''$ et $y \leq y''$. Donc, par définition de R_1, on a $(x, y)R_1(x'', y'')$.
     Donc R_1 est transitive.
   – antisymétrie :
     Soit $(x, y), (x', y') \in \mathbb{N} \times \mathbb{N}$.
     Supposons que $(x, y)R_1(x', y')$ et $(x', y')R_1(x, y)$. Montrons que $(x, y) = (x', y')$.
     Par définition de R_1, on a $x \leq x'$, $y \leq y'$, $x' \leq x$ et $y' \leq y$.
     Comme $\leq$ est antisymétrique, on en déduit que $x = x'$ et $y = y'$. Donc $(x, y) = (x', y')$.
     Donc R_1 est antisymétrique.

   On en déduit que R_1 est une relation d'ordre.
2. R_1 n'est pas une relation d'ordre totale car les couples $(1, 2)$ et $(2, 1)$ sont incomparables.
   En effet, raisonnons par l'absurde et supposons que $(1, 2)R_1(2, 1)$ ou $(2, 1)R_1(1, 2)$.
   – Premier cas : $(1, 2)R_1(2, 1)$.
     Par définition de R_1, cela signifie que $1 \leq 2$ et $2 \leq 1$.
     Par antisymétrie de $\leq$, on obtient donc $1 = 2$, ce qui n'est pas. Contradiction.
   – Second cas : $(2, 1)R_1(1, 2)$.
     Par définition de R_1, cela signifie que $2 \leq 1$ et $1 \leq 2$.
     Par antisymétrie de $\leq$, on obtient $2 = 1$, ce qui n'est pas. Contradiction.

   On a donc montré que $\neg(((1, 2)R_1(2, 1)) \vee ((2, 1)R_1(1, 2)))$, c'est-à-dire que les couples $(1, 2)$ et $(2, 1)$ sont incomparables selon R_1. Donc R_1 n'est pas un ordre total.
3. Montrons que R_2 est une relation d'ordre :
   – réflexivité :
     Soit $(x, y) \in \mathbb{N} \times \mathbb{N}$. Montrons que $(x, y)R_2(x, y)$.
     Comme $\leq$ est réflexive, on a $y \leq y$. De plus $x = x$. Donc $(x, y)R_2(x, y)$.
     Donc R_2 est réflexive.
   – transitivité :
     Soit $(x, y), (x', y'), (x'', y'') \in \mathbb{N} \times \mathbb{N}$.
     Supposons que $(x, y)R_2(x', y')$ et $(x', y')R_2(x'', y'')$. Montrons que $(x, y)R_2(x'', y'')$.
     Par définition de $(x, y)R_2(x', y')$, on a deux cas possibles :

- Soit $x < x'$:
  Par définition de $(x', y')R_2(x'', y'')$, on a deux cas possibles :
  - Soit $x' < x''$:
    Dans ce cas, comme $<$ est transitive, on a $x < x''$ et donc $(x, y)R_2(x'', y'')$.
  - Soit $x' = x''$ et $y' \leq y''$
    On a donc, dans ce cas, $x < x' = x''$, c'est-à-dire $x < x''$ et donc $(x, y)R_2(x'', y'')$.
- Soit $x = x'$ et $y \leq y'$
  Par définition de $(x', y')R_2(x'', y'')$, on a deux cas possibles :
  - Soit $x' < x''$:
    Dans ce cas, on a $x = x' < x''$ et donc $(x, y)R_2(x'', y'')$.
  - Soit $x' = x''$ et $y' \leq y''$
    On a alors $x = x' = x''$. De plus, $y \leq y'$ et $y' \leq y''$ donc, comme $\leq$ est transitive, on a $y \leq y''$. Ainsi, $x = x''$ et $y \leq y''$, donc $(x, y)R_2(x'', y'')$.

Dans tous les cas, on a montré que $(x, y)R_2(x'', y'')$.
Donc R_2 est transitive.
- antisymétrie :
  Soit $(x, y), (x', y') \in \mathbb{N} \times \mathbb{N}$.
  Supposons que $(x, y)R_2(x', y')$ et $(x', y')R_2(x, y)$. Montrons que $(x, y) = (x', y')$.
  En raisonnant par l'absurde, supposons que $(x, y) \neq (x', y')$. On a alors $x \neq x'$ ou $y \neq y'$.
  - Supposons dans un premier temps que $x \neq x'$.
    Par définition de $(x, y)R_2(x', y')$, on a donc $x < x'$.
    De même, par définition de $(x', y')R_2(x, y)$, on a $x' < x$.
    On a donc $x < x'$ et $x' < x$, ce qui est impossible, car par transitivité on obtient $x < x$ et $<$ est irréflexive.
    Contradiction.
  - Maintenant, dans le cas où $x = x'$, on a nécessairement $y \neq y'$.
    Comme $<$ est irréflexive, on a $x \not< x$.
    Par définition de $(x, y)R_2(x', y')$, on a $y \leq y'$ (et $x = x'$).
    De même, par définition de $(x', y')R_2(x, y)$, on a $y' \leq y$ (et $x' = x$).
    On a donc $y \leq y'$ et $y' \leq y$. Par antisymétrie de $\leq$, on obtient $y = y'$, ce qui contredit l'hypothèse.

  Dans les deux cas, on a abouti à une contradiction, donc $(x, y) = (x', y')$.
  Donc R_2 est antisymétrique.

On en déduit que R_2 est une relation d'ordre.

4. Montrons que R_2 est une relation d'ordre totale.
   Soit $(x, y), (x', y') \in \mathbb{N} \times \mathbb{N}$. Montrons que $(x, y)R_2(x', y')$ ou $(x', y')R_2(x, y)$.
   Comme $\leq$ est un ordre total, on a $x \leq x'$ ou $x' \leq x$.
   Par symétrie, on peut supposer par exemple que $x \leq x'$.
   On a alors deux cas :
   - Premier cas : $x = x'$.
     Alors, comme $\leq$ est un ordre total, on a $y \leq y'$ ou $y' \leq y$.
     - Si $y \leq y'$, on a alors, par définition de R_2, $(x, y)R_2(x', y')$ (car $x = x'$ et $y \leq y'$).
     - Si $y' \leq y$, on a alors, par définition de R_2, $(x', y')R_2(x, y)$ (car $x' = x$ et $y' \leq y$).
   - Second cas : $x \neq x'$.
     On a alors $x \leq x'$ et $x \neq x'$, c'est-à-dire $x < x'$. Par définition de R_2, on a donc $(x, y)R_2(x', y')$.

   On a ainsi montré que R_2 est un ordre total.
5. Il est facile de voir que $R_1 \subset R_2$, c'est-à-dire que R_1 est un raffinement de R_2.

Montrons donc que

$$\forall (x,y),(x',y') \in \mathbb{N}\times\mathbb{N} : (x,y)R_1(x',y') \Rightarrow (x,y)R_2(x',y')$$

Soit donc $(x,y),(x',y') \in \mathbb{N}\times\mathbb{N}$.
Supposons que $(x,y)R_1(x',y')$ et montrons que $(x,y)R_2(x',y')$.
Par définition de R_1, on a $x \leq x'$ et $y \leq y'$.
On raisonne par cas suivant que $x = x'$ ou non.

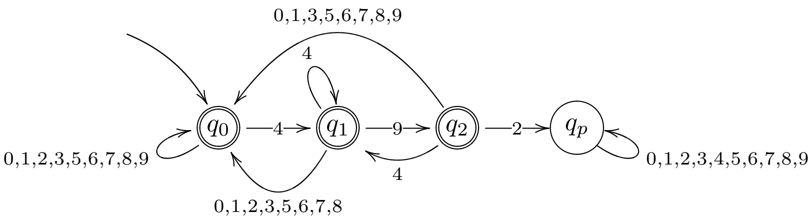
- Premier cas : $x = x'$.
  On a alors $x = x'$ et $y \leq y'$, donc $(x,y)R_2(x',y')$.
- Second cas : $x \neq x'$.
  On a alors $x \leq x'$ et $x \neq x'$, c'est-à-dire $x < x'$. Par définition de R_2, on a donc $(x,y)R_2(x',y')$.

Dans les deux cas, on a montré que $(x,y)R_2(x',y')$.
Donc R_1 est un raffinement de R_2.

■

3. Correction d'automate

Soit $\Sigma \triangleq \{0,1,2,3,4,5,6,7,8,9\}$ un alphabet et $A_{-492} = (Q,\Sigma,\delta,q_0,F)$ l'automate trouvé lors de la série précédente et donné ci-dessous.



Montrer que A_{-492} accepte exactement le langage L_{-492} constitué de l'ensemble des chaînes qui ne contiennent pas le sous-mot 492. Pour cela :

1. Décrire formellement le langage L_{-492} en définissant un ensemble de chaînes sur Σ par compréhension.
2. Donner le tableau de l'automate A_{-492}. Déterminer les conditions nécessaires pour que $\delta(q,a) = q_1$, $\delta(q,a) = q_2$ et $\delta(q,a) = q_p$.
3. En utilisant les conditions du point précédent, démontrer par induction sur la structure des mots (décomposer à droite) la proposition suivante :

$$\forall w \in \Sigma^* : \widehat{\delta}(q_0,w) = q_p \Leftrightarrow \exists x,y \in \Sigma^* : w = x492y$$

4. Déduire de la proposition précédente que $\mathrm{L}(A_{-492}) = L_{-492}$. □

Solution.

1. $L_{-492} = \{w \in \Sigma^* \mid \neg(\exists x,y \in \Sigma^* : w = x492y)\}$
2. Le tableau est :

δ	0	1	2	3	4	5	6	7	8	9
SFq_0	q_0	q_0	q_0	q_0	q_1	q_0	q_0	q_0	q_0	q_0
Fq_1	q_0	q_0	q_0	q_0	q_1	q_0	q_0	q_0	q_0	q_2
Fq_2	q_0	q_0	q_p	q_0	q_1	q_0	q_0	q_0	q_0	q_0
q_p	q_p	q_p	q_p	q_p	q_p	q_p	q_p	q_p	q_p	q_p

et les conditions nécessaires sont les suivantes,

$$\delta(q,a) = q_1 \Leftrightarrow a = 4 \wedge q \neq q_P \tag{D1}$$
$$\delta(q,a) = q_2 \Leftrightarrow q = q_1 \wedge a = 9 \tag{D2}$$
$$\delta(q,a) = q_p \Leftrightarrow (q = q_p) \vee (q = q_2 \wedge a = 2) \tag{D3}$$

3. On démontre par induction sur la structure des mots la proposition suivante :

$$\underbrace{\forall w \in \Sigma^* : \widehat{\delta}(q_0, w) = q_p \Leftrightarrow \underbrace{\exists x, y \in \Sigma^* : w = x492y}_{Q(w)}}_{P(w)}$$

(a) Cas de base $P(\epsilon)$.
(⇒) Supposons $\widehat{\delta}(q_0, \epsilon) = q_p$, nous avons par définition de $\widehat{\delta}$, $\widehat{\delta}(q_0, \epsilon) = q_0 \neq q_p$, contradiction. On déduit donc $Q(w)$.
(⇐) Supposons que $\exists x, y \in \Sigma^* : \epsilon = x492y$, nous avons $|\epsilon| = 0 \neq 3 \leq |x492y|$, ce qui signifie que $\forall x, y \in \Sigma^* : \epsilon \neq x492y$, contradiction. On déduit donc $\widehat{\delta}(q_0, \epsilon) = q_p$

(b) Induction $P(w) \Rightarrow P(w \cdot a)$ avec $a \in \Sigma$.
On suppose $P(w)$. Nous avons

$$\begin{aligned} \widehat{\delta}(q_0, wa) = q_p &\Leftrightarrow \delta(\widehat{\delta}(q_0, w), a) = q_p && \text{def. de } \widehat{\delta} \\ &\Leftrightarrow \big(\widehat{\delta}(q_0, w) = q_p\big) \vee \big(\widehat{\delta}(q_0, w) = q_2 \wedge a = 2\big) && \text{par (D3)} \end{aligned}$$

Supposons que la dernière proposition ci-dessus est logiquement équivalente à $Q(wa)$ alors nous avons $P(wa)$. Reste donc à montrer que cette proposition est équivalente $Q(wa)$.
(⇒) On démontre que chacun des membres de la disjonction implique $Q(wa)$.

i. $$\begin{aligned} \widehat{\delta}(q_0, w) = p_q &\Rightarrow \exists x, y \in \Sigma^* : w = x492y && \text{hyp. d'induct.} \\ &\Rightarrow \exists x', y' \in \Sigma^* : wa = x'492y' && (x' = x \text{ et } y' = ya) \\ &\Leftrightarrow Q(wa) \end{aligned}$$

ii. L'état initial étant q_0, on a $\widehat{\delta}(q_0, w) = q_2$ ssi $|w| \geq 2$. Dès lors on peut décomposer $w = w'cb$ avec $w' \in \Sigma^*$ et $c, b \in \Sigma$. Par ailleurs nous avons,

$$\begin{aligned} \widehat{\delta}(q_0, w) = q_2 \Leftrightarrow \widehat{\delta}&(q_0, w'cb) \\ &\Leftrightarrow \delta(\widehat{\delta}(q_0, w'c), b) = q_2 && \text{def. de } \widehat{\delta} \\ &\Leftrightarrow \widehat{\delta}(q_0, w'c) = q_1 \wedge b = 9 && \text{par (D2)} \\ &\Leftrightarrow \delta(\widehat{\delta}(q_0, w'), c) = q_1 \wedge b = 9 && \text{def. de } \widehat{\delta} \\ &\Leftrightarrow c = 4 \wedge b = 9 && \text{par (D1)} \end{aligned}$$

et donc

$$\begin{aligned} \widehat{\delta}(q_0, w) = q_2 \wedge a = 2 &\Leftrightarrow \exists w' \in \Sigma^* : w = w'49 \\ &\Rightarrow \exists x, y \in \Sigma^* : wa = x492y \quad (x = w' \text{ et } y = \epsilon) \\ &\Leftrightarrow Q(wa) \end{aligned}$$

($\Leftarrow$) On démontre que $Q(wa)$ implique toujours un des membres de la disjonction. Supposons $Q(wa)$:

$$\exists x, y \in \Sigma^* : wa = x492y$$

Il y a deux cas à considérer,

i. $y \neq \epsilon$. Dans ce cas, en posant $x' = x$ et y' tel que $y = y'a$, on a $\exists x', y' \in \Sigma^* : w = x'492y'$, c'est à dire $Q(w)$, et donc $\widehat{\delta}(q_0, w) = p_q$ par l'hypothèse d'induction.

ii. $y = \epsilon$. On a donc $a = 2$ et $\exists x \in \Sigma^* : w = x49$. Deux cas se présentent. Si x contient le mot 492 alors w satisfait $Q(w)$ et par induction on a $\widehat{\delta}(q_0, w) = p_q$.
Si x ne contient pas le mot 492, alors on a $\neg Q(w)$ et l'on sait par l'hypothèse d'induction que$^{(**)}$ $\widehat{\delta}(q_0, w) \neq q_p$

$$\begin{aligned}\widehat{\delta}(q_0, w) &= \widehat{\delta}(q_0, x49)\\ &= \delta(\widehat{(}q_0, x4), 9)\\ &= \delta(\delta(\widehat{\delta}(q_0, x), 4), 9)\end{aligned}$$

sachant $(**)$, on sait que $\widehat{\delta}(q_0, x) \neq q_p$ et par (D1) on a donc

$$\begin{aligned}&= \delta(q_1, 9) && \text{(D1)}\\ &= q_2 && \text{par def. de } \delta\end{aligned}$$

on a donc $\widehat{\delta}(q_0, w) = q_2 \wedge a = 2$, l'autre cas de la disjonction.

Par (a) et (b) et le principe d'induction sur les mots nous avons $\forall w \in \Sigma^* : P(w)$.

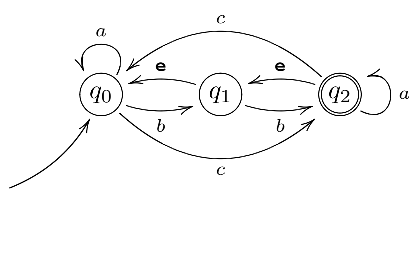
4. Pour l'automate A_{-492} nous avons$^{(*)}$ $q \in F \Leftrightarrow q \neq q_p$.

$$\begin{aligned}\mathrm{L}(A_{-492}) &= \left\{w \in \Sigma^* \mid \widehat{\delta}(q_0, w) \in F\right\} && \text{def. de } \mathrm{L}(A_{-492})\\ &= \left\{w \in \Sigma^* \mid \widehat{\delta}(q_0, w) \neq q_p\right\} && \text{par } (*)\\ &= \left\{w \in \Sigma^* \mid \neg Q(w)\right\} && \text{proposition 3.}\\ &= \left\{w \in \Sigma^* \mid \neg\left(\exists x, y \in \Sigma^* : w = x492y\right)\right\} &&\\ &= L_{-492} && \text{déf. de } L_{-492}\end{aligned}$$

■

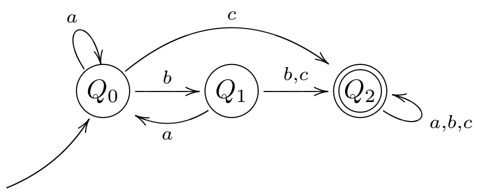
4. Élimination des transitions instantanées

Soit $\Sigma \triangleq \{a, b, c\}$. Déterminiser l'AFN$_{\mathbf{e}}$ M sur Σ donné ci-dessous, c'est à dire calculer $\mathrm{D}_{\mathbf{e}}(M)$ et donner le graphe de l'automate obtenu.



□

Solution. Dans la joie et la bonne humeur, on obtient :



avec $Q_0 \triangleq \{q_0\}$, $Q_1 \triangleq \{q_0, q_1\}$, $Q_2 \triangleq \{q_0, q_1, q_2\}$. ∎

5. Concaténation de langages

Soit Σ un alphabet et deux AFD $M = (Q, \Sigma, \delta, s, F)$ et $M' = (Q', \Sigma, \delta', s', F')$ tels que $Q \cap Q' = \emptyset$. Définissez de manière formelle un AFN$_{\mathbf{e}}$ M'' tel que :

$$\mathrm{L}(M'') = \mathrm{L}(M) \cdot \mathrm{L}(M')$$

Pourquoi avons nous supposé que $Q \cap Q' = \emptyset$? □

Solution. L'idée est d'ajouter une transition instantanée des états finaux de l'automate M dans l'état de départ de M'. L'état initial est celui de M et les états finaux sont ceux de l'automate M'. On définit donc l'AFN$_{\mathbf{e}}$ par :

$$M'' = (Q \cup Q', \Sigma, \delta'', s, F')$$

avec

$$\delta''(q, a) = \begin{cases} \{\delta(q, a)\} & \text{si } q \in Q \wedge a \neq \mathbf{e} \\ \{\delta'(q, a)\} & \text{si } q \in Q' \wedge a \neq \mathbf{e} \\ \emptyset & \text{si } q \in \big((Q \cup Q') \setminus F\big) \wedge a = \mathbf{e} \\ \{s'\} & \text{si } q \in F \wedge a = \mathbf{e} \end{cases}$$

Ci-dessus, le troisième cas s'explique par le fait que δ'' devant être totale sur $\Sigma \cup \{\mathbf{e}\}$, il faut rajouter une transition instantanée pour *chacun* des états de $Q \cup Q'$.

Nous supposons que $Q \cap Q' = \emptyset$ afin de ne pas mélanger les états du premier AFD avec ceux du second. Sans cette hypothèse il est possible, dans certain cas, que δ'' ne soit pas une fonction et donc M'' ne soit pas un AFN$_{\mathbf{e}}$.

Notons que cette contrainte n'est en fait pas restrictive. Il est toujours possible de renommer l'ensemble des états Q_1 d'un automate et de changer δ, F et s de façon adéquate pour se retrouver avec un nouvel automate qui accepte le même langage mais avec un ensemble Q_2 d'états tel que $Q_1 \cap Q_2 = \emptyset$. Ces deux automates sont dits *isomorphes*. ∎