Grundlagen der Theoretischen Informatik

Till Mossakowski

Fakultät für Informatik
Otto-von-Guericke Universität
Magdeburg

Wintersemester 2014/15

Äquivalenz von NEA und DEA

Zwei endliche Automaten M_1 und M_2 heißen *äquivalent* genau dann wenn $L(M_1) = L(M_2)$.

Satz: [Rabin, Scott]
Zu jedem nichtdeterministischen endlichen Automaten gibt es einen äquivalenten deterministischen endlichen Automaten.

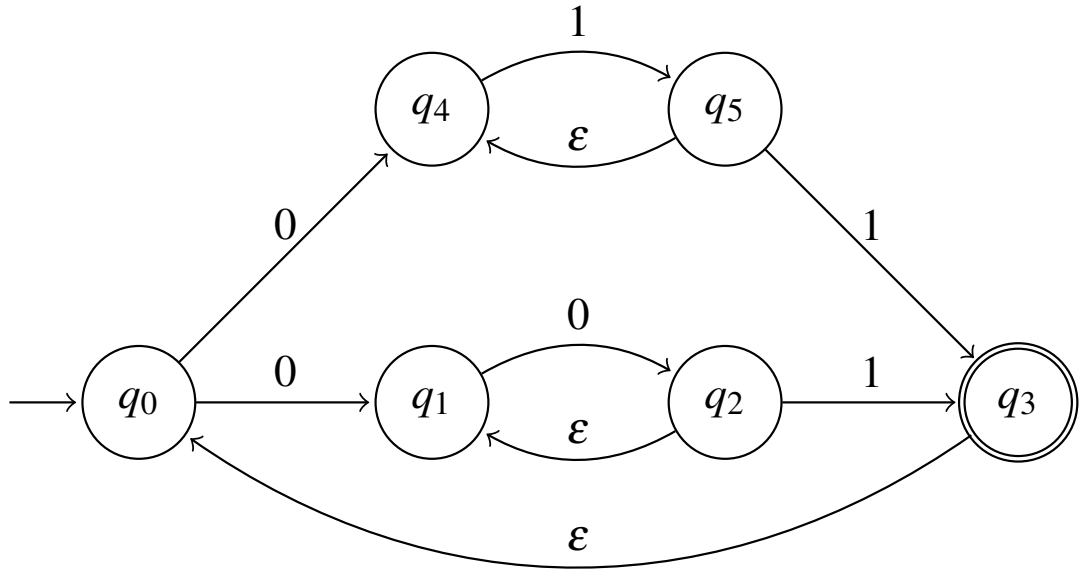
Beweisidee:

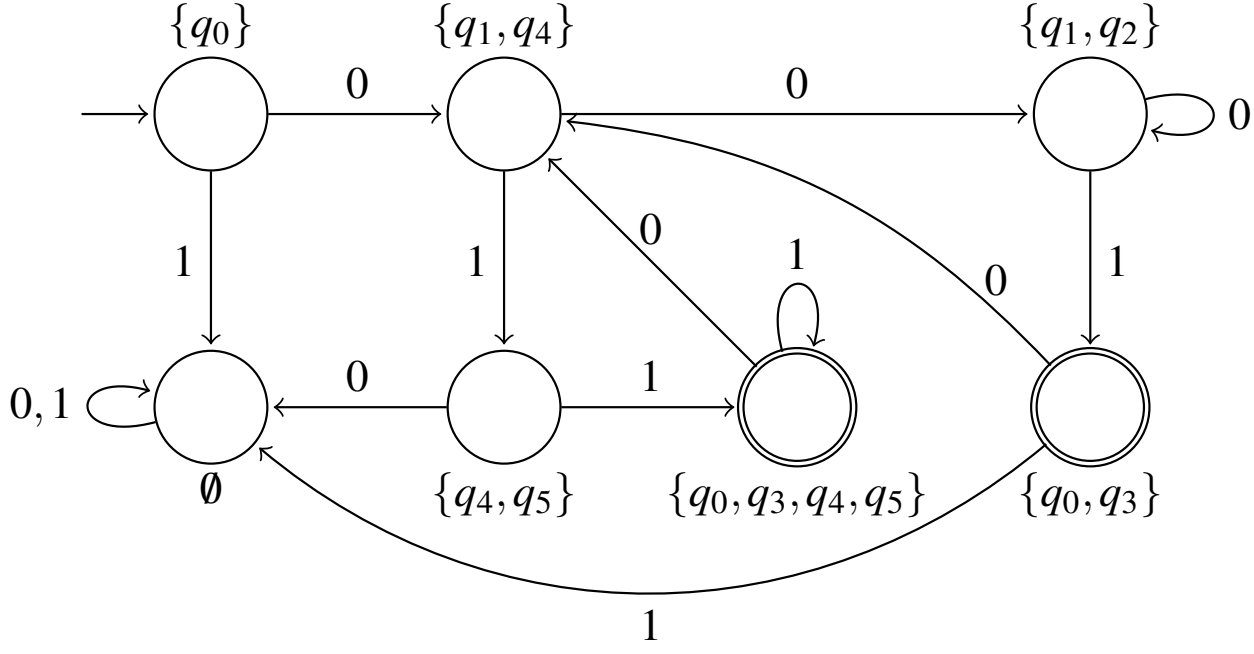
Zustände des simulierenden deterministischen endlichen Automaten $M' = (K', \Sigma, \delta', .., .)$ sind Mengen von Zuständen des simulierten nichtdeterministischen endlichen Automaten $M = (K, \Sigma, \Delta, s, F)$.

$$K' = 2^K$$

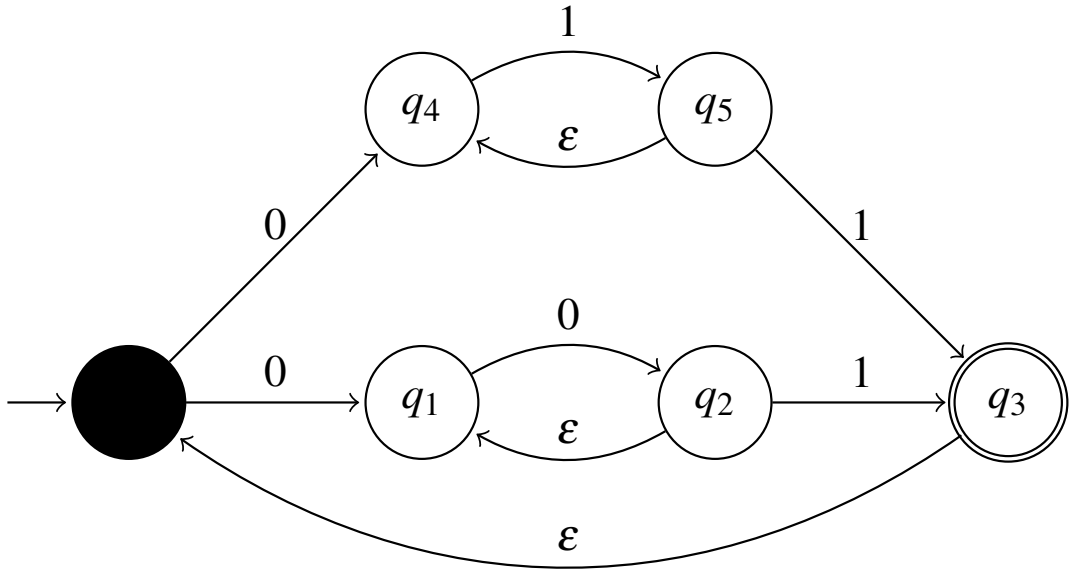
M' verfolgt alle möglichen Berechnungen von M „synchron“:

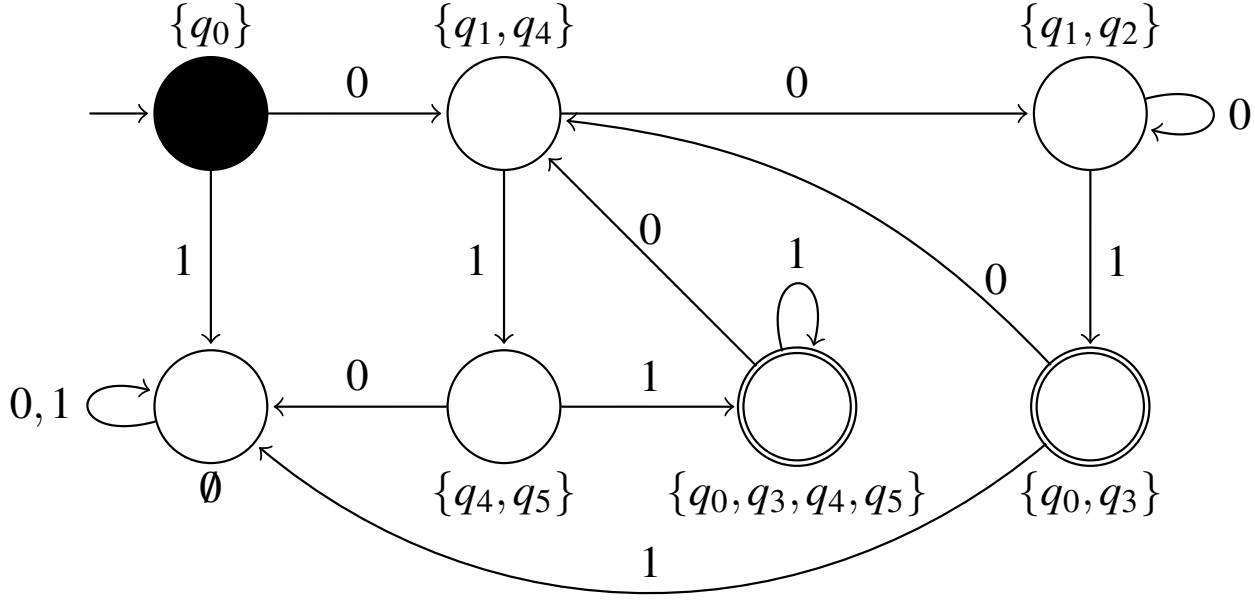
$$q \in Q \subseteq K \quad \wedge \quad (q, a, p) \in \Delta \quad \Rightarrow \quad p \in \delta'(Q, a)$$


1
q_4
q_5
ε
0
1
0
0
q_0
1
q_1
q_2
q_3
ε
ε

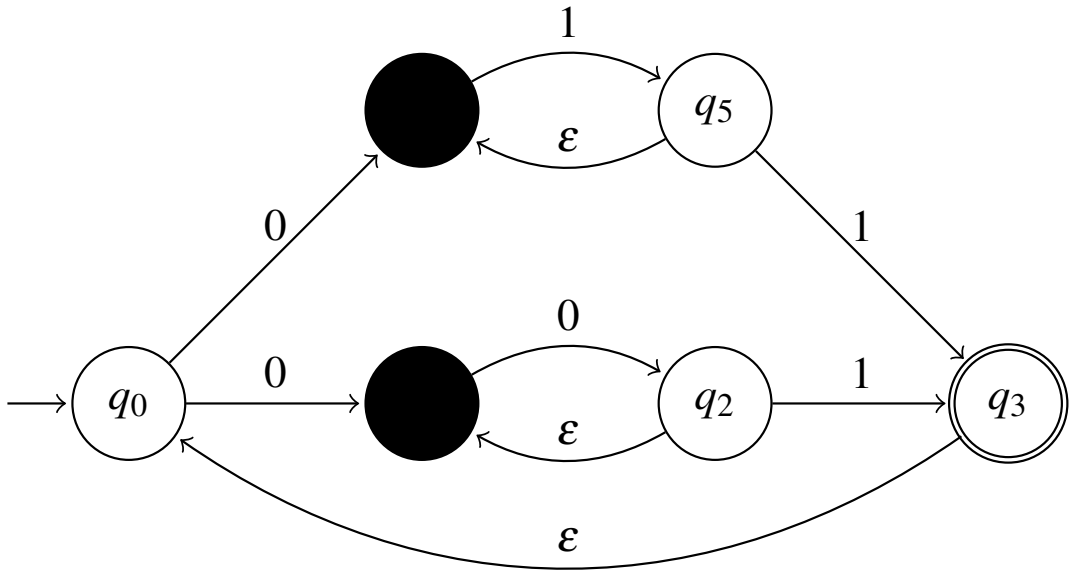

$\{q_0\}$
$\{q_1, q_4\}$
$\{q_1, q_2\}$
0
0
0
1
1
1
0
1
0
0, 1
0
1
$\emptyset$
$\{q_4, q_5\}$
$\{q_0, q_3, q_4, q_5\}$
$\{q_0, q_3\}$
1

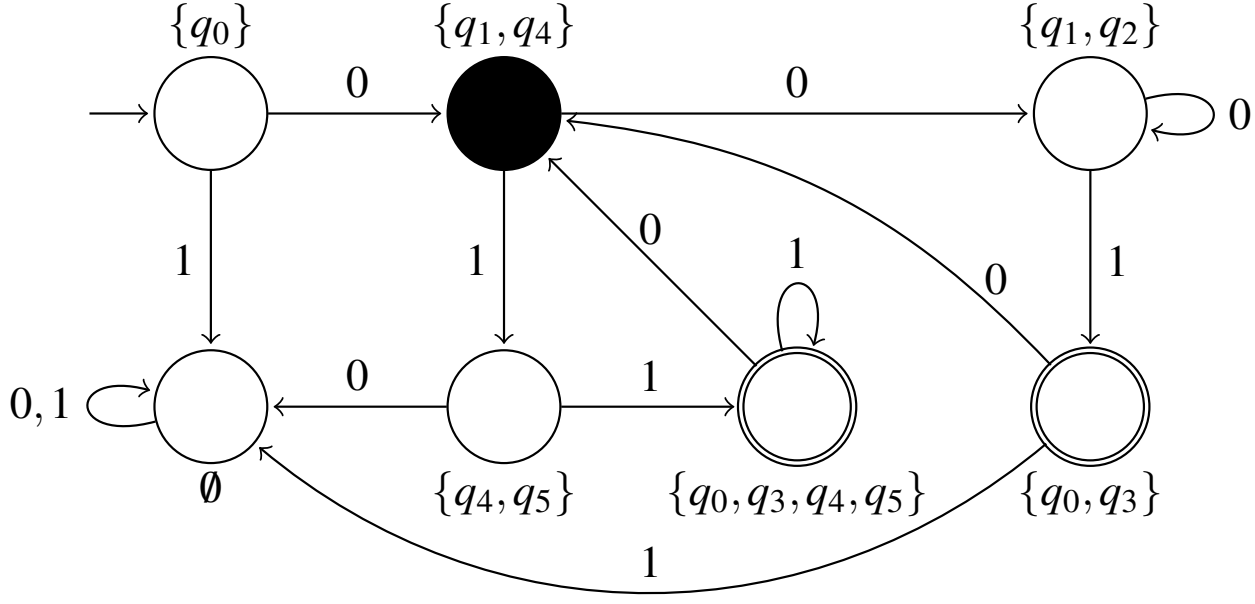
011001


1
q_4
q_5
ε
0
1
0
0
q_1
q_2
1
q_3
ε
ε

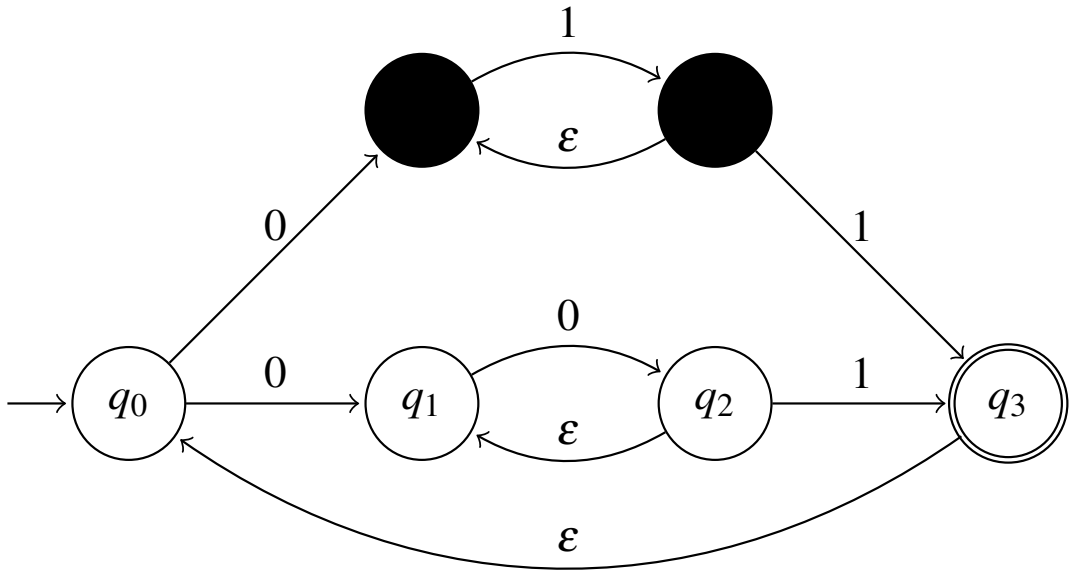

$\{q_0\}$
$\{q_1, q_4\}$
$\{q_1, q_2\}$
0
0
0
1
1
0
1
0
1
0,1
0
1
$\emptyset$
$\{q_4, q_5\}$
$\{q_0, q_3, q_4, q_5\}$
$\{q_0, q_3\}$
1

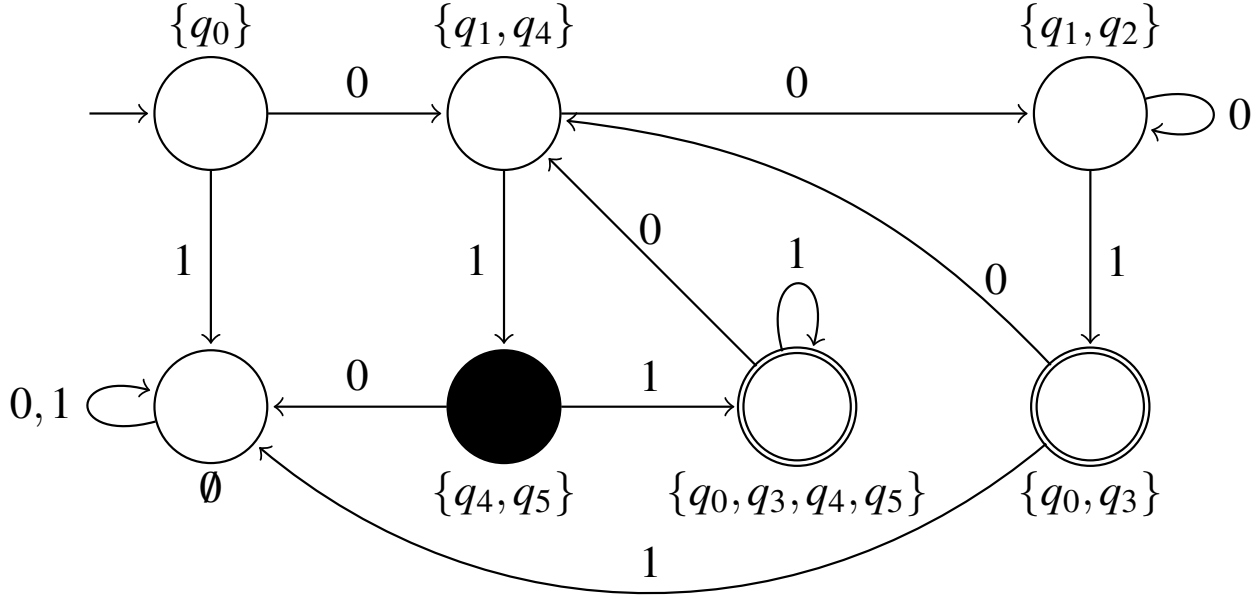
011001


1
q_5
ε
0
1
0
0
q_0
0
q_2
1
q_3
ε
ε

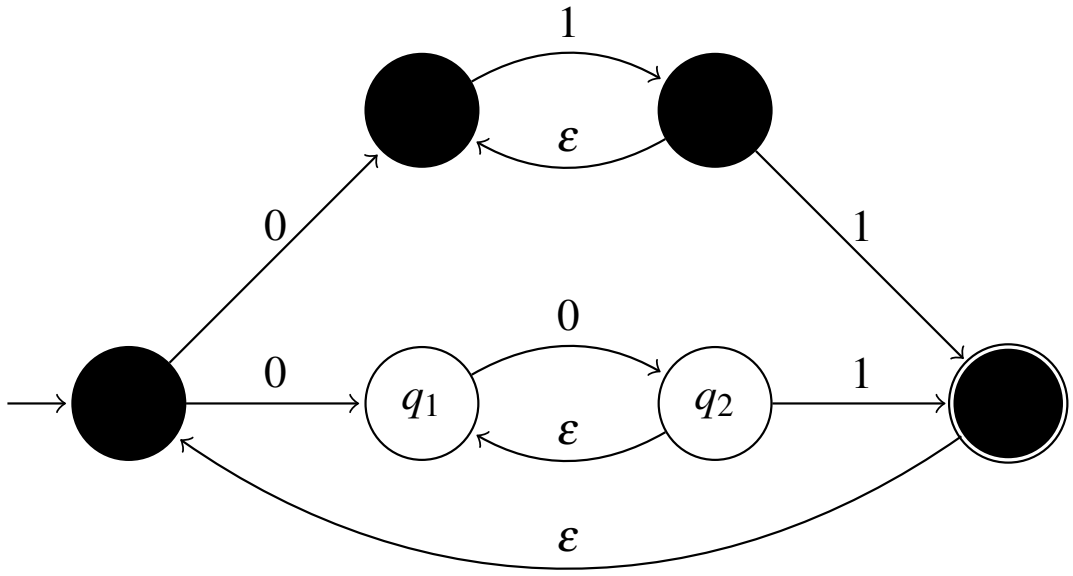

$\{q_0\}$
$\{q_1, q_4\}$
$\{q_1, q_2\}$
0
0
0
1
1
0
1
0
1
0,1
0
1
$\emptyset$
$\{q_4, q_5\}$
$\{q_0, q_3, q_4, q_5\}$
$\{q_0, q_3\}$
1

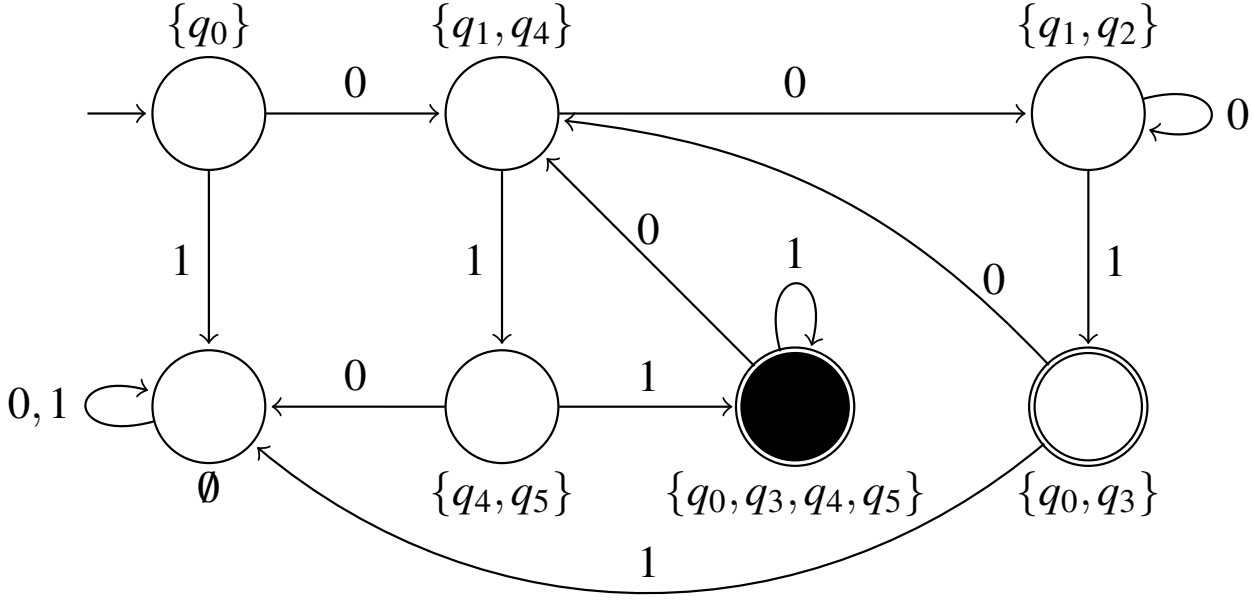
011001


1
ε
0
1
0
0
1
q_0
q_1
q_2
q_3
ε
ε

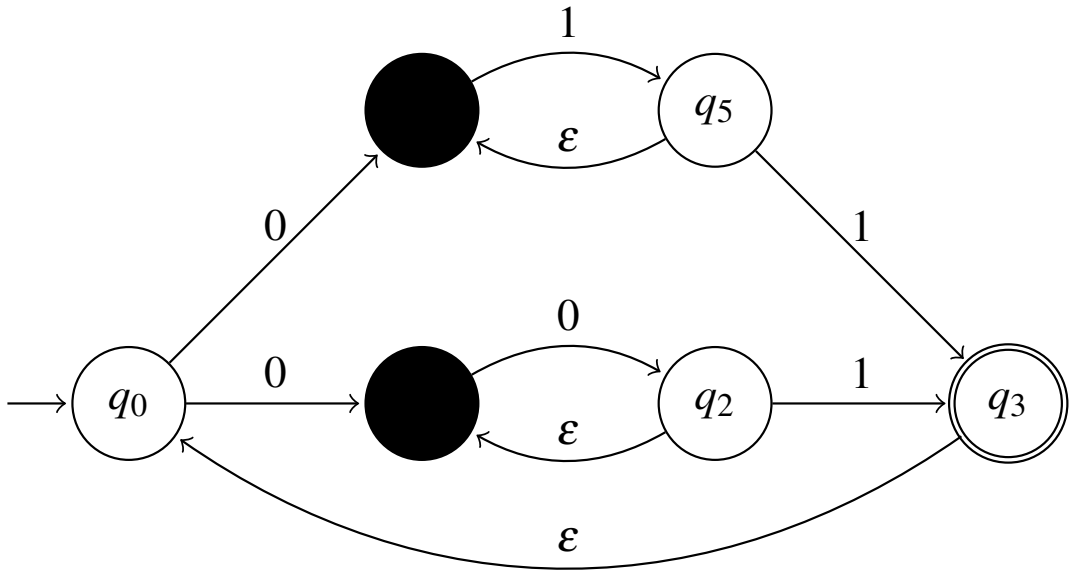

$\{q_0\}$
$\{q_1, q_4\}$
$\{q_1, q_2\}$
0
0
0
1
1
1
0
1
0
0, 1
0
1
$\emptyset$
$\{q_4, q_5\}$
$\{q_0, q_3, q_4, q_5\}$
$\{q_0, q_3\}$
1

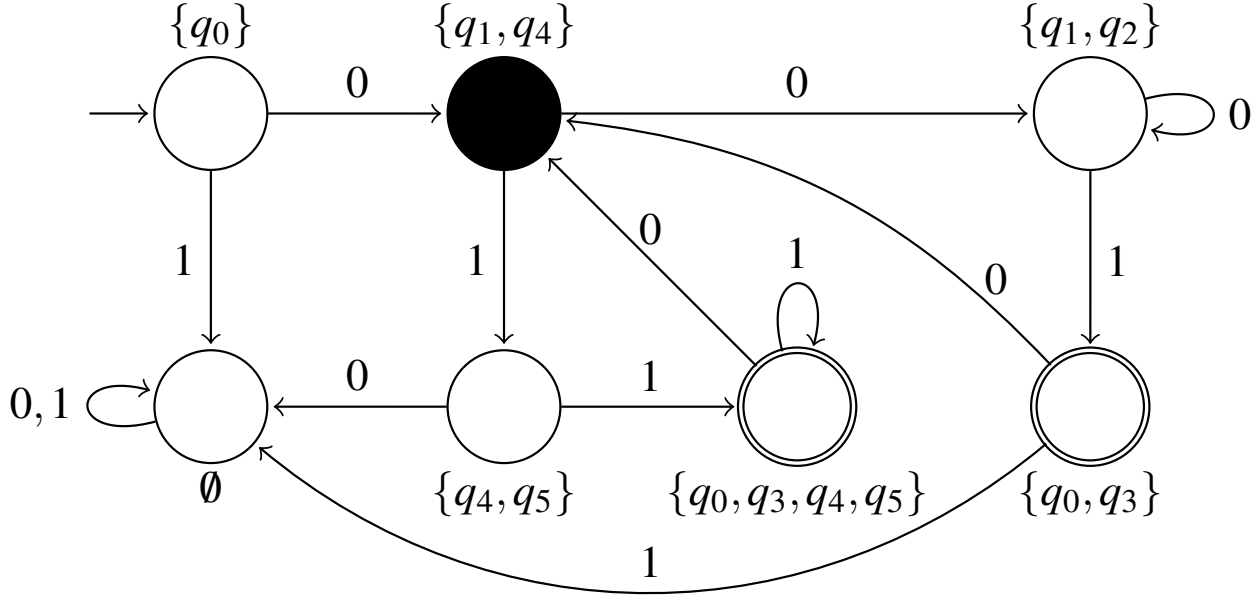
011001


1
ε
0
1
0
0
q1
ε
q2
1
ε

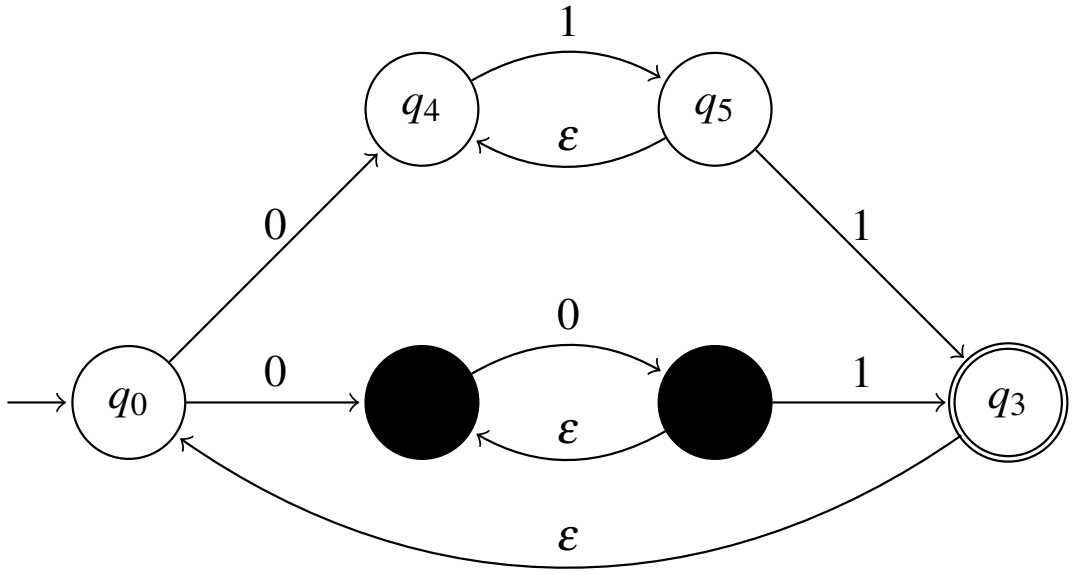

{q0}
{q1, q4}
{q1, q2}
0
0
0
1
1
0
1
1
0
0, 1
0
1
∅
{q4, q5}
{q0, q3, q4, q5}
{q0, q3}
1

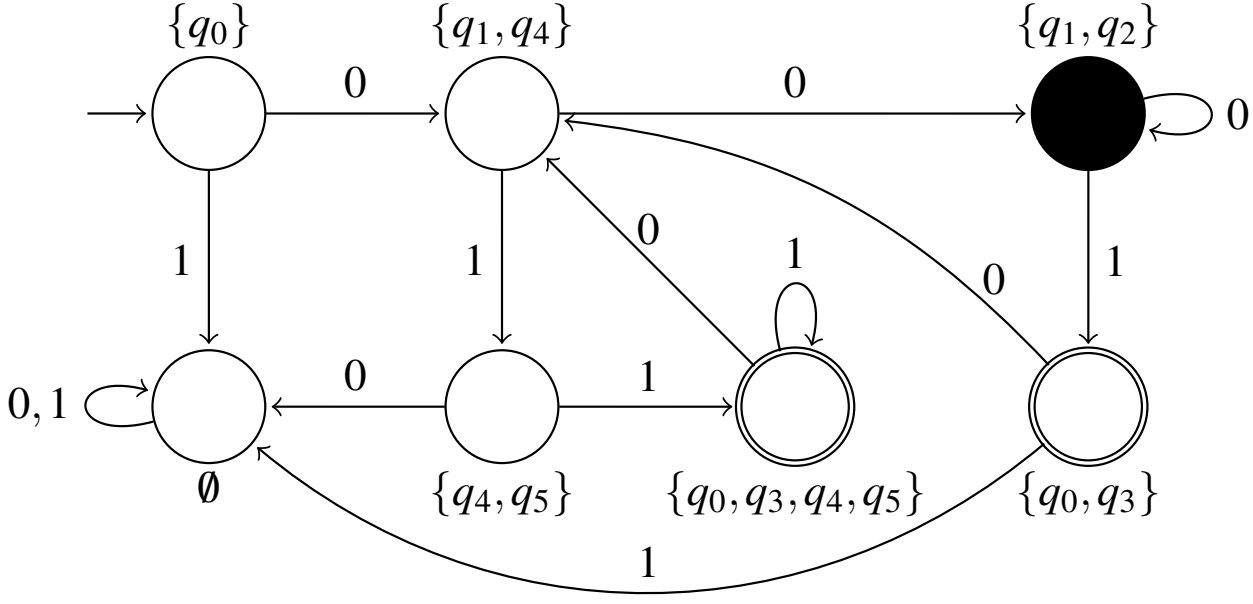
011001


1
q_5
ε
0
1
0
0
q_0
0
q_2
1
q_3
ε
ε

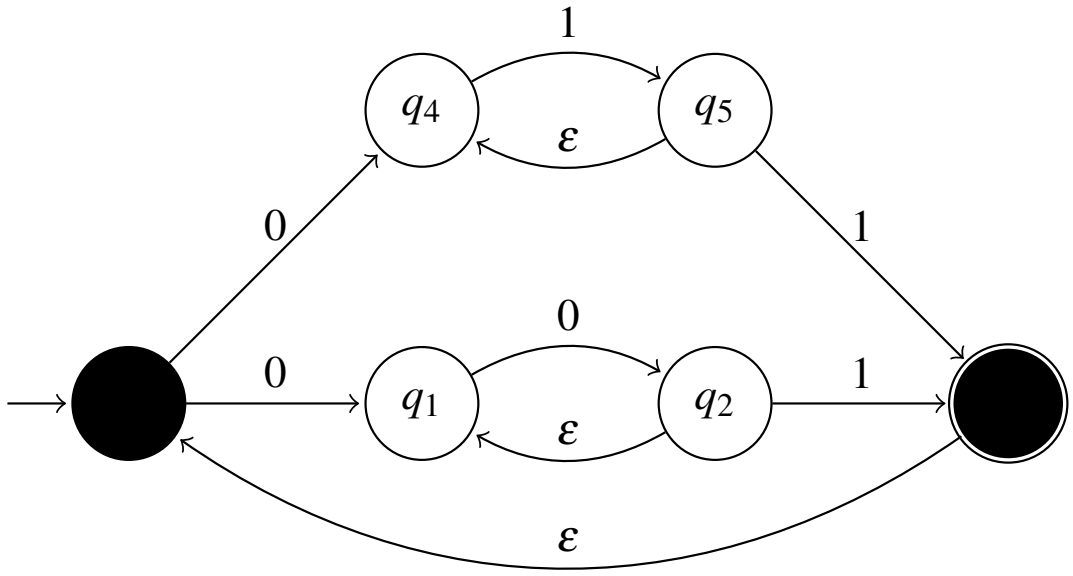

$\{q_0\}$
$\{q_1, q_4\}$
$\{q_1, q_2\}$
0
0
0
1
1
0
1
0
1
0,1
0
1
$\emptyset$
$\{q_4, q_5\}$
$\{q_0, q_3, q_4, q_5\}$
$\{q_0, q_3\}$
1

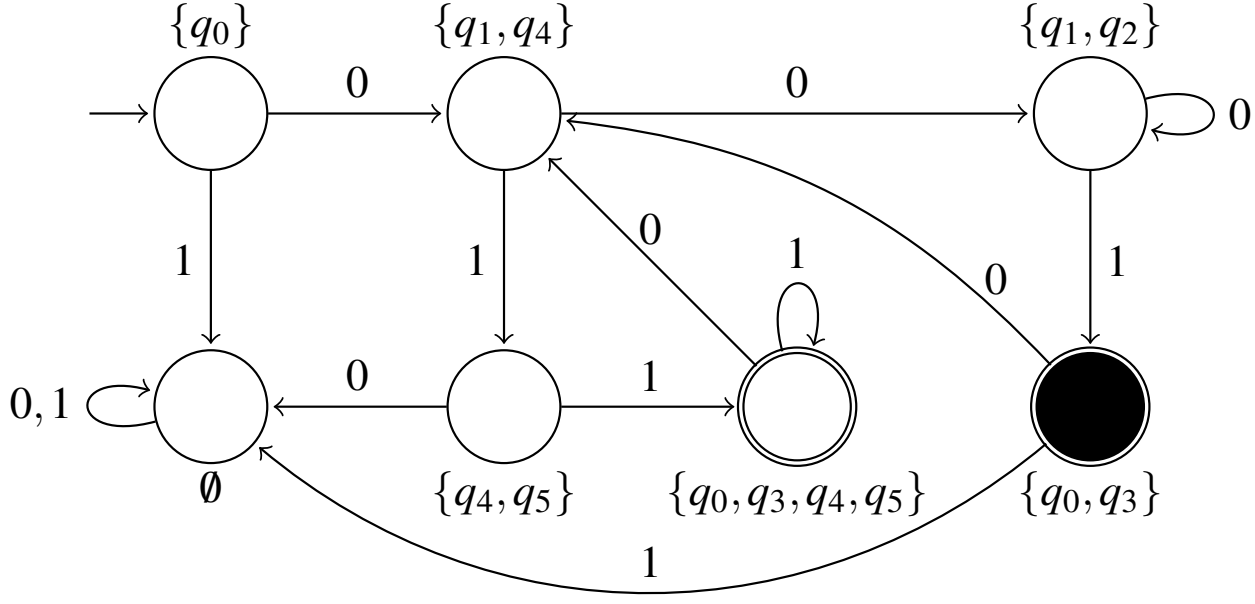
011001


1
q_4
q_5
ε
0
1
0
q_0
0
1
q_3
ε
ε


$\{q_0\}$
$\{q_1, q_4\}$
$\{q_1, q_2\}$
0
0
0
1
1
0
1
0
1
0,1
0
1
$\emptyset$
$\{q_4, q_5\}$
$\{q_0, q_3, q_4, q_5\}$
$\{q_0, q_3\}$
1

011001


1
q_4
q_5
ε
0
1
0
0
q_1
q_2
1
ε
ε


$\{q_0\}$
$\{q_1, q_4\}$
$\{q_1, q_2\}$
0
0
0
1
1
0
1
0
1
0, 1
0
1
$\emptyset$
$\{q_4, q_5\}$
$\{q_0, q_3, q_4, q_5\}$
$\{q_0, q_3\}$
1

Bezüglich eines nichtdeterministischen endlichen Automaten $M = (K, \Sigma, \Delta, s, F)$ und $q \in K$ sei $E(q) = \{p \in K \mid (q, \varepsilon) \vdash_M^* (p, \varepsilon)\}$ der *ε-Abschluss* von q.

Beweis: Sei $M = (K, \Sigma, \Delta, s, F)$ ein nichtdeterministischer endlicher Automat. Sei $M' = (K', \Sigma, \delta', s', F')$ wobei

$$\begin{aligned} K' &= 2^K \\ s' &= E(s) \\ F' &= \{Q \subseteq K \mid Q \cap F \neq \emptyset\} \\ \delta'(Q, a) &= \bigcup_{p \in K \mid \exists q \in Q : (q,a,p) \in \Delta} E(p) \end{aligned}$$

Wir müssen zeigen, dass $w \in L(M) \Leftrightarrow w \in L(M')$ oder anders gesagt, wir müssen zeigen, dass gilt:

$$\exists f \in F : (s,w) \vdash_M^* (f,\varepsilon) \Leftrightarrow \exists Q \in F' : (s',w) \vdash_{M'}^* (Q,\varepsilon)$$

Es genügt dazu, die folgende stärkere Behauptung zu zeigen:

$$\begin{aligned} (q,w) \vdash_M^* (p,\varepsilon) &\Rightarrow \exists P : p \in P \wedge (E(q),w) \vdash_{M'}^* (P,\varepsilon) \\ (E(q),w) \vdash_{M'}^* (P,\varepsilon) &\Rightarrow \forall p \in P : (q,w) \vdash_M^* (p,\varepsilon) \end{aligned}$$

Wir beweisen die Behauptung durch Induktion über $|w|$.

Bei der Induktionsverankerung ist $w = \varepsilon$ und es ist zu zeigen:

$$\begin{aligned} (q,\varepsilon) \vdash_M^* (p,\varepsilon) &\Rightarrow \exists P : p \in P \wedge (E(q),\varepsilon) \vdash_{M'}^* (P,\varepsilon) \\ (E(q),\varepsilon) \vdash_{M'}^* (P,\varepsilon) &\Rightarrow \forall p \in P : (q,\varepsilon) \vdash_M^* (p,\varepsilon) \end{aligned}$$

Der erste Teil gilt mit $P = E(q)$ nach Definition von $E(q)$.
Beim zweiten Teil ist notwendigerweise $P = E(q)$, da M' ein deterministischer Automat ist, und die Behauptung folgt aus der Definition von $E(q)$.

Für den Induktionsschritt nehmen wir nun an, dass die Behauptung für alle Wörter der Länge höchstens k gilt.

Sei nun $w = va$ mit $|v| = k$ und $a \in \Sigma$. Dann ist zu zeigen

$$\begin{aligned} (q, va) \vdash_M^* (p, \varepsilon) &\Rightarrow \exists P : p \in P \wedge (E(q), va) \vdash_{M'}^* (P, \varepsilon) \\ (E(q), va) \vdash_{M'}^* (P, \varepsilon) &\Rightarrow \forall p \in P : (q, va) \vdash_M^* (p, \varepsilon) \end{aligned}$$

Falls $(q, va) \vdash_M^* (p, \varepsilon)$ gilt, so gibt es Zustände r_1 und r_2, so dass $(q, va) \vdash_M^* (r_1, a) \vdash_M (r_2, \varepsilon) \vdash_M^* (p, \varepsilon)$.

Also gilt auch $(q, v) \vdash_M^* (r_1, \varepsilon)$ und nach Induktionsannahme gibt es ein R mit $r_1 \in R$ so dass $(E(q), v) \vdash_{M'}^* (R, \varepsilon)$ und somit gilt auch $(E(q), va) \vdash_{M'}^* (R, a)$.

Nach Konstruktion von M' gilt wegen $(r_1, a, r_2) \in \Delta$

$$E(r_2) \subseteq \delta'(R, a)$$

Wegen $(r_2, \varepsilon) \vdash_M^* (p, \varepsilon)$ gilt $p \in E(r_2)$ und nach Definition von δ' gilt $p \in \delta'(R, a)$.

Teil 1 der Behauptung folgt somit mit $P = \delta'(R, a)$:

$$(E(q), va) \vdash_{M'}^* (R, a) \vdash_{M'} (P, \varepsilon)$$

Betrachten wir Teil 2 der Behauptung. Sei R der Zustand in K', der vor P erreicht wird:

$$(E(q), va) \vdash_{M'}^{*} (R, a) \vdash_{M'} (P, \varepsilon)$$

Dann ist $\delta'(R, a) = P$. Nach Definition von δ' gibt es für alle $p \in \delta'(R, a)$ ein $r_1 \in R$ und $r_2 \in K$ mit $(r_1, a, r_2) \in \Delta$ und $p \in E(r_2)$.

Nach Induktionsannahme angewandt auf $(E(q), v) \vdash_{M'}^{*} (R, \varepsilon)$ gilt für alle $r \in R$, dass $(q, v) \vdash_{M}^{*} (r, \varepsilon)$.

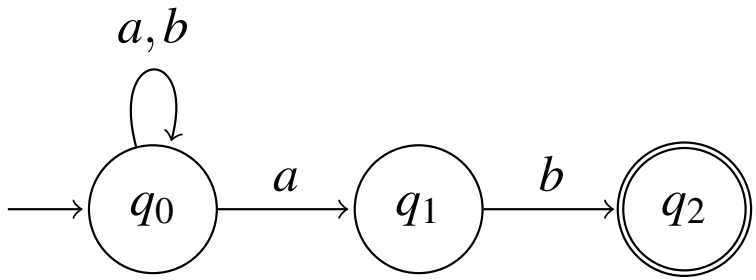
Somit gibt es für alle $p \in P = \delta'(R, a)$ ein $r_1 \in R$ und ein $r_2 \in K$ mit

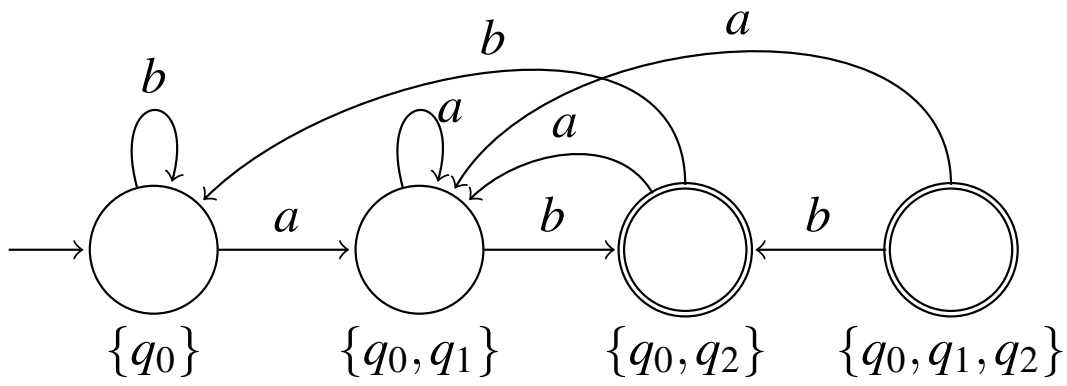
$$(q, va) \vdash_{M}^{*} (r_1, a) \vdash_{M} (r_2, \varepsilon) \vdash_{M}^{*} (p, \varepsilon)$$

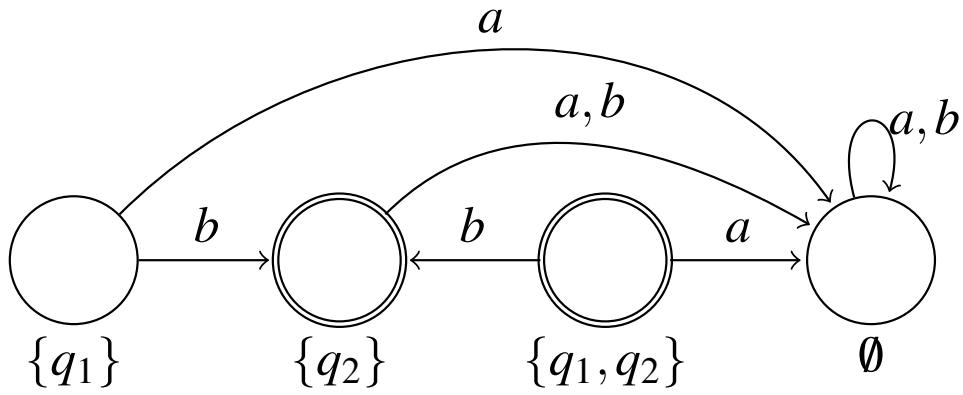
Damit sind Behauptung und Satz bewiesen. ■

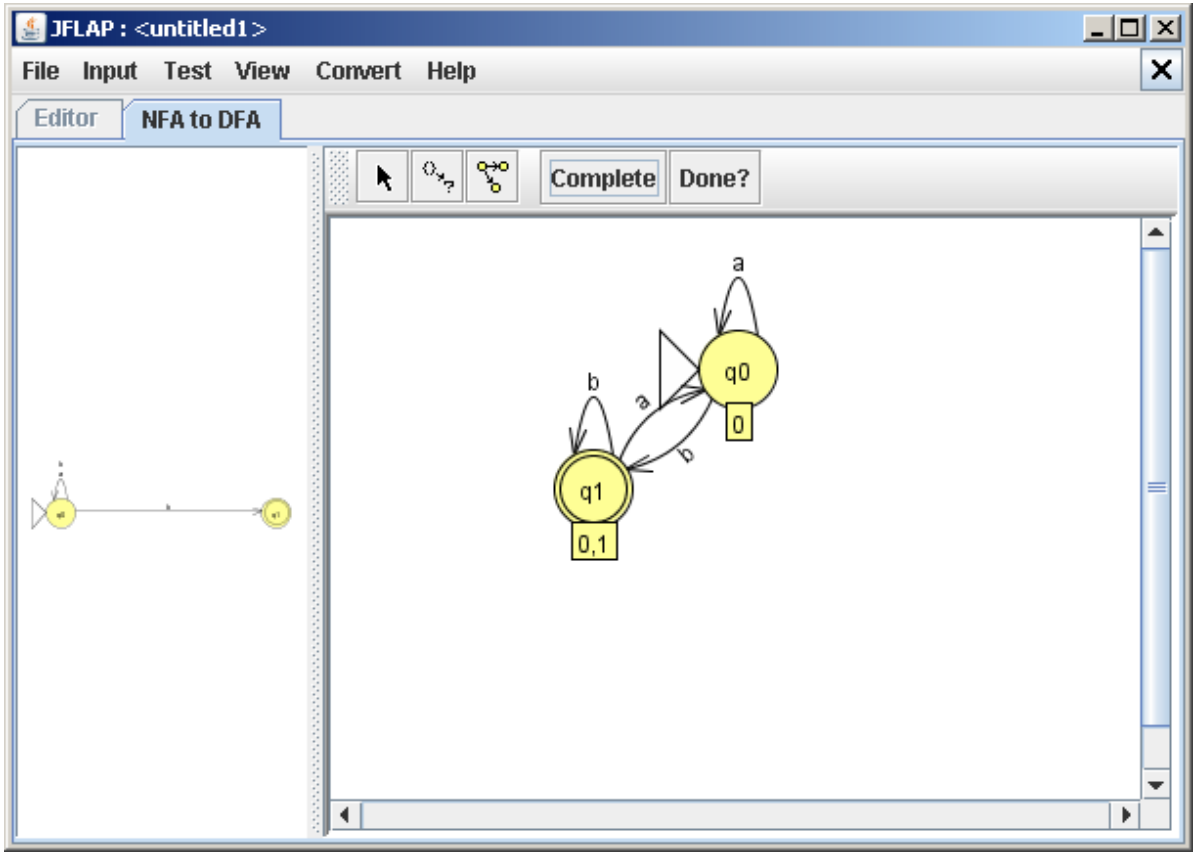
Oft sind einige der Zustände des im Beweis konstruierten DEA nicht relevant, weil sie vom Startzustand aus nicht erreichbar sind.

Beispiel:




b
b
a
a
a
b
b
a
$\{q_0\}$
$\{q_0, q_1\}$
$\{q_0, q_2\}$
$\{q_0, q_1, q_2\}$


a
a, b
a, b
b
b
a
$\{q_1\}$
$\{q_2\}$
$\{q_1, q_2\}$
$\emptyset$


JFLAP : <untitled1>
File
Input
Test
View
Convert
Help
Editor
NFA to DFA
Complete
Done?
a
q0
0
b
a
b
q1
0,1

Abschlusseigenschaften

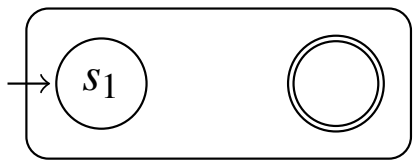
Eine Menge M heißt *abgeschlossen* unter einer Operation, falls die Operation angewandt auf Elemente aus M stets wieder Elemente aus M ergibt.

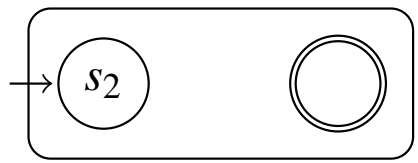
Satz:
Die Klasse der von endlichen Automaten akzeptierten Sprachen ist abgeschlossen unter

(a) Vereinigung,
(b) Konkatenation,
(c) Kleene Star,
(d) Komplement und
(e) Schnitt.

Beweisskizze:

Seien $M_1 = (K_1, \Sigma, \Delta_1, s_1, F_1)$ und $M_2 = (K_2, \Sigma, \Delta_2, s_2, F_2)$ nichtdeterministische endliche Automaten mit (o.B.d.A.) $K_1 \cap K_2 = \emptyset$ und $s \notin K_1 \cup K_2$.

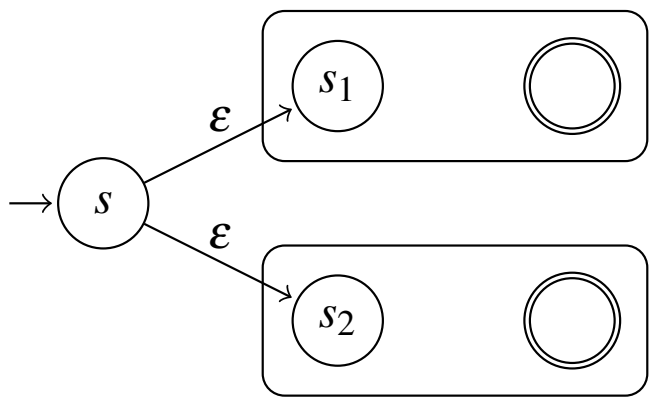




(a) Vereinigung

Sei $M = (K, \Sigma, \Delta, s, F)$ und

$$\begin{aligned} K &= K_1 \cup K_2 \cup \{s\} \\ F &= F_1 \cup F_2 \\ \Delta &= \Delta_1 \cup \Delta_2 \cup \{(s, \varepsilon, s_1), (s, \varepsilon, s_2)\} \end{aligned}$$

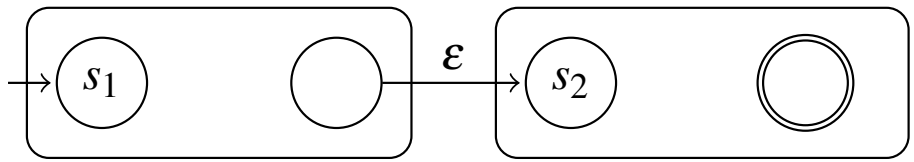


Dann gilt $L(M) = L(M_1) \cup L(M_2)$

(b) Konkatenation

Sei $M = (K, \Sigma, \Delta, s_1, F)$ mit

$$\begin{aligned} K &= K_1 \cup K_2 \\ F &= F_2 \\ \Delta &= \Delta_1 \cup \Delta_2 \cup \{(f, \varepsilon, s_2) \mid f \in F_1\} \end{aligned}$$

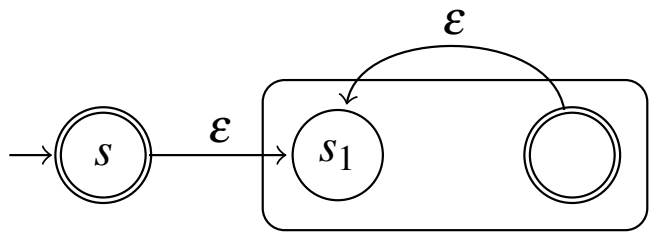


Dann gilt $L(M) = L(M_1)L(M_2)$

(c) Kleene Star

Sei $M = (K, \Sigma, \Delta, s, F)$ und

$$\begin{aligned} K &= K_1 \cup \{s\} \\ F &= F_1 \cup \{s\} \\ \Delta &= \Delta_1 \cup \{(s, \varepsilon, s_1)\} \cup \{(f, \varepsilon, s_1) \mid f \in F_1\} \end{aligned}$$



Dann gilt $L(M) = L(M_1)^*$

(d) Komplement

Sei $M_1 = (K_1, \Sigma, \delta_1, s_1, F_1)$ ein deterministischer endlicher Automat.

Sei $M = (K_1, \Sigma, \delta_1, s_1, F)$ mit $F = K_1 - F_1$.

Dann gilt $L(M) = \Sigma^* - L(M_1)$

(e) Schnitt

Seien $M_1 = (K_1, \Sigma, \delta_1, s_1, F_1)$ und $M_2 = (K_2, \Sigma, \delta_2, s_2, F_2)$ deterministische endliche Automaten.

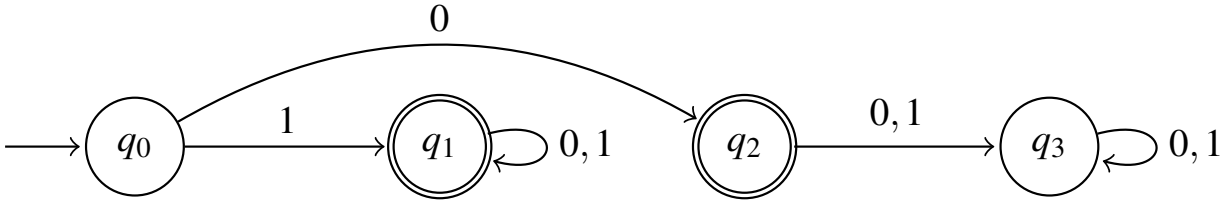
Sei $M = (K, \Sigma, \delta, s, F)$ mit

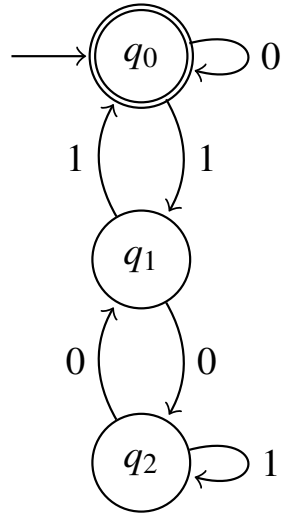
$$\begin{aligned} K &= K_1 \times K_2 \\ s &= (s_1, s_2) \\ F &= F_1 \times F_2 \end{aligned}$$

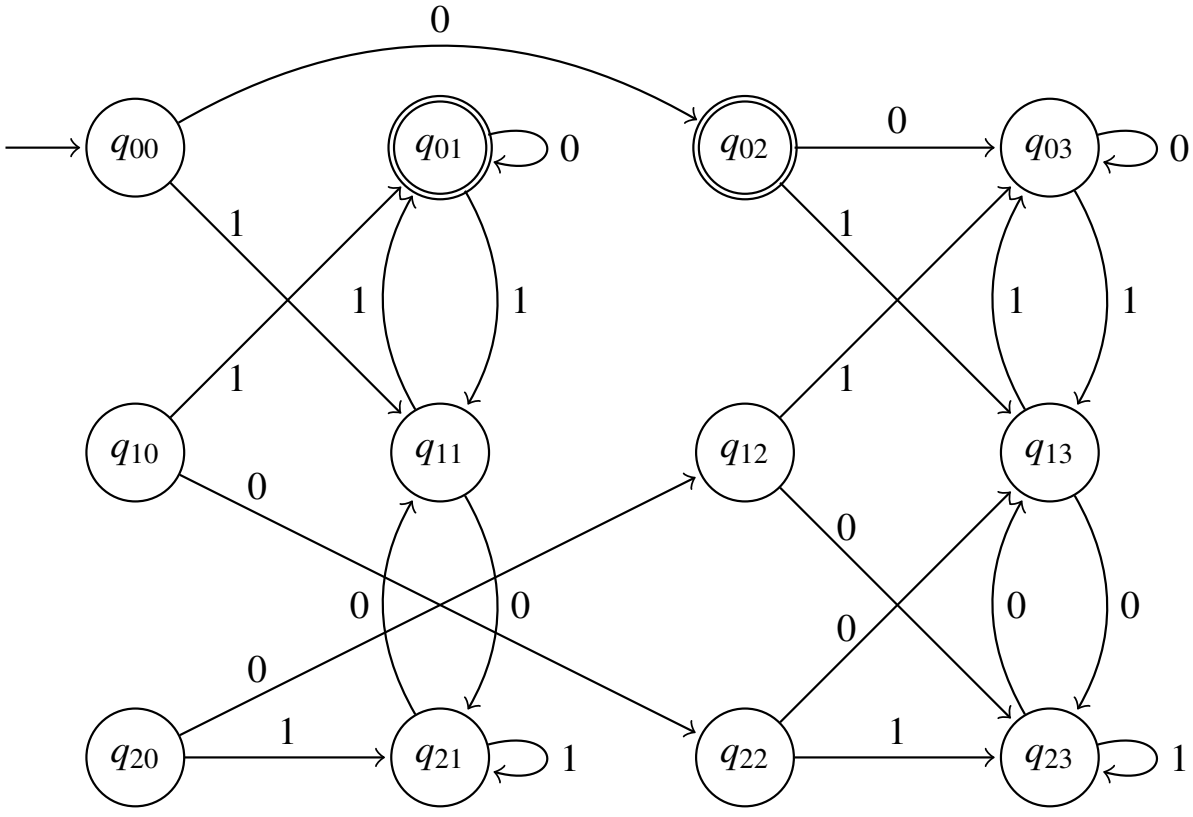
und $\delta((q_1, q_2), a) = (\delta_1(q_1, a), \delta_2(q_2, a))$.

Dann gilt $L(M) = L(M_1) \cap L(M_2)$ ■

Beispiel:


q_0
q_1
q_2
q_3
0
1
0, 1
0, 1
0, 1


q_0
q_1
q_2
0
1
1
0
0
1


q_{00}
q_{01}
q_{02}
q_{03}
q_{10}
q_{11}
q_{12}
q_{13}
q_{20}
q_{21}
q_{22}
q_{23}
0
1

Reguläre Ausdrücke

Wir können Sprachen auch mit Hilfe von Operationen auf Sprachen, insbesondere Konkatenation und Kleene Star, beschreiben.

Beispiele: Sei $\Sigma = \{a, b, c\}$

$\{w \in \Sigma^* \mid w \text{ beginnt mit } ab\}$

$$\{a\} \circ \{b\} \circ \Sigma^*$$

$\{w \in \Sigma^* \mid w \text{ enthält genau ein } b\}$

$$\{a, c\}^* \circ \{b\} \circ \{a, c\}^*$$

Definition:
Sei Σ ein Alphabet mit $\{(,),\emptyset,\cup,^*\} \cap \Sigma = \emptyset$.
Ein regulärer Ausdruck über einem Alphabet Σ ist ein Wort über dem Alphabet $\Sigma \cup \{(,),\emptyset,\cup,^*\}$, das wie folgt erzeugt werden kann:

(1) *Jedes Element von Σ ist ein regulärer Ausdruck.*

(2) *$\emptyset$ ist ein regulärer Ausdruck.*

(3) *Falls α und β reguläre Ausdrücke sind, dann ist $(\alpha\beta)$ ein regulärer Ausdruck.*

(4) *Falls α und β reguläre Ausdrücke sind, dann ist $(\alpha \cup \beta)$ ein regulärer Ausdruck.*

(5) *Falls α ein regulärer Ausdruck ist, dann ist α^* ein regulärer Ausdruck.*

Jeder reguläre Ausdruck repräsentiert eine Sprache:
Sei $\mathcal{L}(\alpha)$ die durch den Ausdruck α repräsentierte Sprache.
$\mathcal{L}$ ist wie folgt definiert:

(1) $\mathcal{L}(a) = \{a\}$ für alle $a \in \Sigma$.

(2) $\mathcal{L}(\emptyset) = \emptyset$.

(3) Falls α und β reguläre Ausdrücke sind, dann ist $\mathcal{L}((\alpha\beta)) = \mathcal{L}(\alpha)\mathcal{L}(\beta)$.

(4) Falls α und β reguläre Ausdrücke sind, dann ist $\mathcal{L}((\alpha \cup \beta)) = \mathcal{L}(\alpha) \cup \mathcal{L}(\beta)$.

(5) Falls α ein regulärer Ausdruck ist, dann ist $\mathcal{L}(\alpha^*) = \mathcal{L}(\alpha)^*$.

Beispiel:

$$\begin{aligned}
\mathcal{L}(((a \cup b)^* a)) &= \mathcal{L}((a \cup b)^*)\mathcal{L}(a) \\
&= \mathcal{L}((a \cup b)^*)\{a\} \\
&= \mathcal{L}((a \cup b))^*\{a\} \\
&= (\mathcal{L}(a) \cup \mathcal{L}(b))^*\{a\} \\
&= (\{a\} \cup \{b\})^*\{a\} \\
&= \{a,b\}^*\{a\}
\end{aligned}$$

$$\mathcal{L}(((a \cup b)^* a)) = \{w \in \{a,b\}^* \mid w \text{ endet mit } a\}$$

Beispiel:

$$\begin{aligned}
\mathcal{L}((((a \cup b)^* b)(a \cup b))) &= \mathcal{L}(((a \cup b)^* b))\mathcal{L}((a \cup b)) \\
&= \mathcal{L}(((a \cup b)^* b))(\mathcal{L}(a) \cup \mathcal{L}(b)) \\
&= \mathcal{L}(((a \cup b)^* b))\{a,b\} \\
&= \mathcal{L}((a \cup b)^*)\mathcal{L}(b)\{a,b\} \\
&= \mathcal{L}((a \cup b))^*\{b\}\{a,b\} \\
&= \{a,b\}^*\{b\}\{a,b\}
\end{aligned}$$

$$\{w \in \{a,b\}^* \mid \text{das vorletzte Symbol in } w \text{ ist ein } b\}$$

Beispiel:

$$\mathcal{L}(\emptyset^*) = \mathcal{L}(\emptyset)^* = \emptyset^* = \{\varepsilon\}$$

Solange der Ausdruck eindeutig bleibt, können wir Klammern weglassen (insbesondere die äußeren Klammern). Zudem können zwecks Lesbarkeit Klammern hinzugefügt werden. Dazu vereinbaren wir, dass Kleene Star höchste Priorität und Konkatenation Vorrang vor Vereinigung hat.

$$\begin{aligned}
\alpha\beta\gamma &\equiv (\alpha\beta)\gamma \\
\alpha \cup \beta \cup \gamma &\equiv (\alpha \cup \beta) \cup \gamma \\
\alpha \cup \beta\gamma &\equiv \alpha \cup (\beta\gamma) \\
\alpha \cup \beta^* &\equiv \alpha \cup (\beta^*) \\
\alpha\beta^* &\equiv \alpha(\beta^*) \\
\alpha\beta\gamma^* &\equiv (\alpha\beta)(\gamma^*)
\end{aligned}$$

Beispiele: $\Sigma = \{a, b\}$

$\{w \in \Sigma^* \mid w \text{ enthält genau ein } a\}$

$$b^* a b^*$$

$\{w \in \Sigma^* \mid w \text{ enthält mindestens ein } a\}$

$$(a \cup b)^* a (a \cup b)^*$$

$\{w \in \Sigma^* \mid \text{auf jedes } a \text{ folgt in } w \text{ mindestens ein } b\}$

$$(b \cup ab)^*$$

Beispiele: $\Sigma = \{a, b\}$

$\{w \in \Sigma^* \mid w$ hat gerade Länge$\}$

$$((a \cup b)(a \cup b))^*$$

$\{w \in \Sigma^* \mid w$ hat ungerade Länge$\}$

$$(a \cup b)((a \cup b)(a \cup b))^*$$

$\{w \in \Sigma^* \mid w$ beginnt und endet mit dem gleichen Symbol$\}$

$$a(a \cup b)^* a \cup b(a \cup b)^* b \cup a \cup b$$

Reguläre Ausdrücke und reguläre Sprachen

Satz: [Kleene]
Die Klasse der durch reguläre Ausdrücke beschreibbaren Sprachen ist genau die Klasse der regulären Sprachen.

Beweis: Dass es für jede Sprache, die durch einen regulären Ausdruck beschrieben werden kann, stets einen endlichen Automaten gibt, der die Sprache akzeptiert, folgt aus der Beobachtung, dass $\emptyset$ und $\{a\}$ für beliebiges $a \in \Sigma$ von endlichen Automaten akzeptierte Sprachen sind und dem Satz über die Abschlusseigenschaften der von endlichen Automaten akzeptierten Sprachen.

Wir müssen also noch zeigen, dass es für jede Sprache L, die von einem endlichen Automaten akzeptiert wird, einen regulären Ausdruck α gibt mit $\mathcal{L}(\alpha) = L$.

Sei $M = (K, \Sigma, \Delta, s, F)$ ein endlicher Automat mit $K = \{q_1, \ldots, q_n\}$ und sei $s = q_1$. Für $i, j = 1, \ldots, n$ und $k = 0, 1, \ldots, n$ definieren wir

$$
\begin{aligned}
R(i,j,k) \quad = \quad & \{w \in \Sigma^* \mid M \text{ geht unter } w \text{ von } q_i \text{ nach } q_j \\
& \text{über, d.h., } (q_i, w) \vdash_M^* (q_j, \varepsilon), \text{ ohne dass} \\
& \text{einer der Zwischenzustände ein Zustand} \\
& \text{in } \{q_{k+1}, \ldots, q_n\} \text{ ist. Anfangs- und} \\
& \text{Schlusszustand der „Berechnung“} \\
& \text{zählen hierbei nicht als Zwischenzustände.}\}
\end{aligned}
$$

Wir zeigen durch Induktion über k, dass alle $R(i,j,k)$ durch einen regulären Ausdruck beschreibbar sind.

Induktionsverankerung: Sei $k = 0$. Dann gilt

$$i \neq j \quad : \qquad R(i,j,0) = \{a \in \Sigma \cup \{\varepsilon\} \mid (q_i, a, q_j) \in \Delta\}$$

$$i = j \quad : \qquad R(i,j,0) = \{a \in \Sigma \cup \{\varepsilon\} \mid (q_i, a, q_j) \in \Delta\} \cup \{\varepsilon\}$$

$R(i,j,0)$ ist also für alle $i,j = 1,\ldots,n$ endlich und somit eine durch einen regulären Ausdruck beschreibbare Sprache!

Induktionsschritt: Nehmen wir also an, dass die $R(i,j,l)$ durch einen regulären Ausdruck beschreibbar sind für $l = 0,\ldots,k$ und alle $i,j = 1,\ldots,n$. Da

$$R(i,j,k+1) = R(i,j,k) \cup R(i,k+1,k)R(k+1,k+1,k)^* R(k+1,j,k)$$

ist jedes $R(i,j,k+1)$ durch einen regulären Ausdruck beschreibbar.

Der Satz folgt nun aus

$$L(M) = \bigcup_{j \mid q_j \in F} R(1,j,n)$$

■