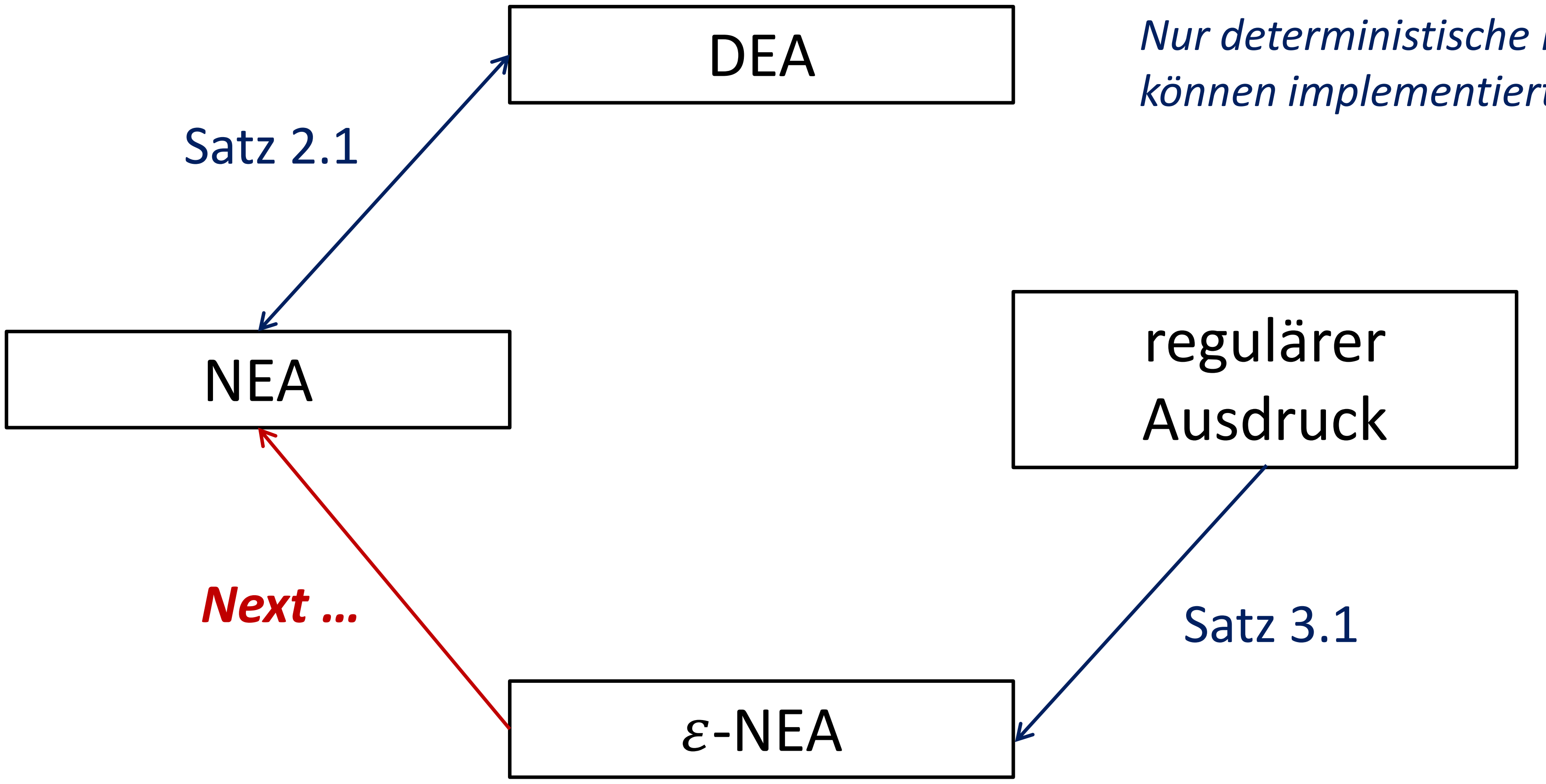
Formale Grundlagen der Informatik

4 Reguläre Ausdrücke und Sprachen

ε-NEA

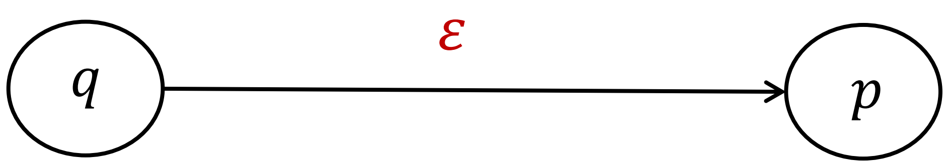
Die vier Mechanismen …



Nur deterministische Modelle können implementiert werden.

Recap: NEA mit ε-Übergängen

- Konnten zu jedem regulären Ausdruck einen ε-NEA konstruieren.



$$p \in \delta(q, \varepsilon)$$

entspricht „spontanem Zustandswechsel" von p nach q

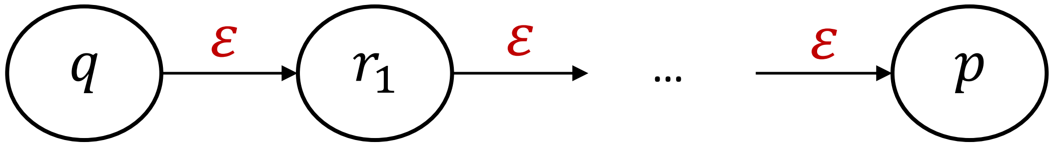
Recap: ε-NEA - Definition

- **Definition:**
  Ein **nichtdeterministischer endlicher Automat mit ε-Übergängen (ε-NEA)** ist ein NEA $A = (Q, \Sigma, \delta, q_0, F)$, dessen Überführungsfunktion zu $\boldsymbol{\delta}: \left(\boldsymbol{Q} \times (\boldsymbol{\Sigma} \cup \{\boldsymbol{\varepsilon}\})\right) \to \mathbf{2}^{\boldsymbol{Q}}$ erweitert ist.

- $\hat{\delta}(q, \varepsilon)$ kann andere Zustände als q enthalten
  - Welche Zustände sind von einem Zustand (nur) mit Hilfe von ε-Übergängen erreichbar?!
  - **ε-Hülle** eines Zustands

ε-Hülle

- Sei $A = (Q, \Sigma, \delta, q_0, F)$ ein ε-NEA und $q \in Q$. Die **$\boldsymbol{\varepsilon}$-Hülle von $\boldsymbol{q}$** ist die Menge aller Zustände p, für die es Zustände $r_0, r_1, \dots, r_k$ gibt, $k \geq 0$, so dass
  1. $r_0 = q$,
  2. $r_k = p$,
  3. $r_{i+1} \in \delta(r_i, \varepsilon)$ für alle i, $0 \leq i < k$.



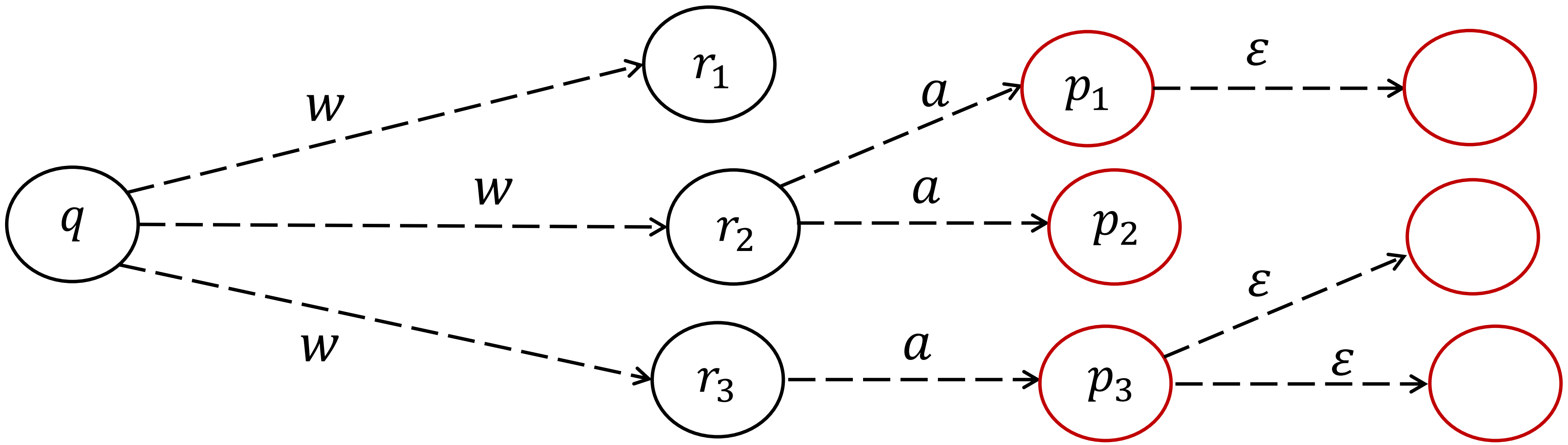
- Sie wird mit **$\boldsymbol{\varepsilon}$H$(\boldsymbol{q})$** bezeichnet.

➢ Wegen der Option $k = 0$ gilt für jeden Zustand q, dass $q \in \varepsilon\mathrm{H}(q)$.

ε-NEA – Erweiterte Überführungsfunktion

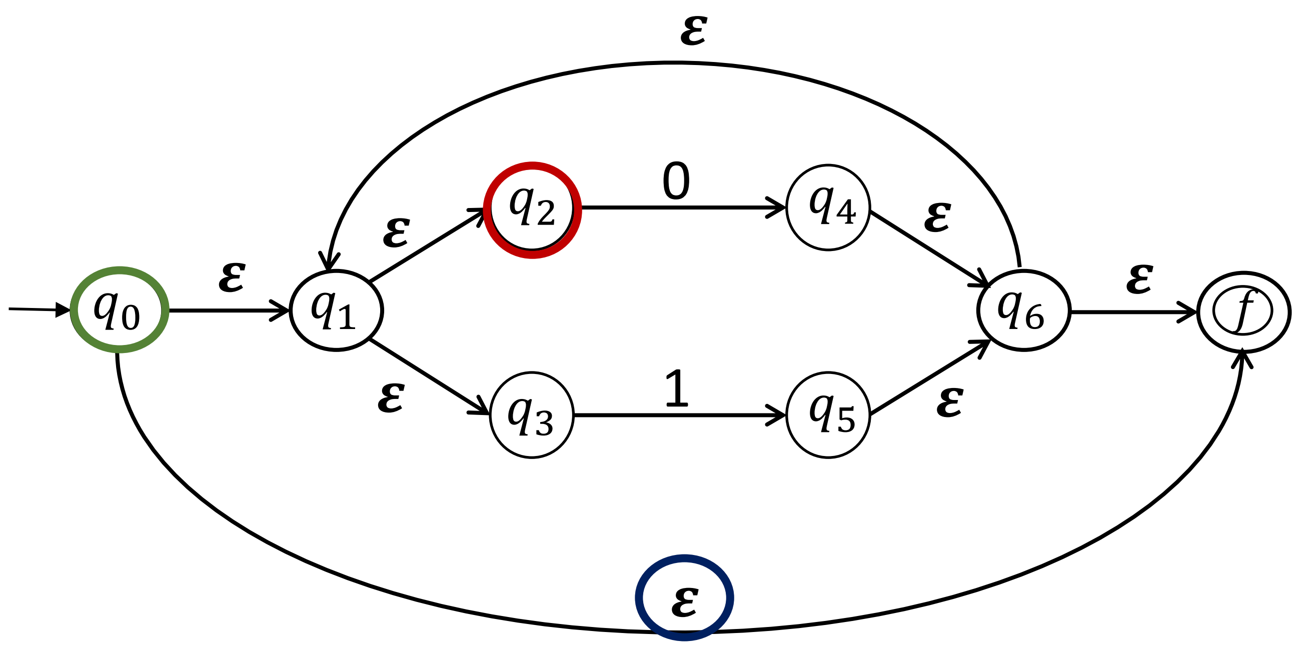
- Sei $A = (Q, \Sigma, \delta, q_0, F)$ ein ε-NEA und $q \in Q, w \in \Sigma^*, a \in \Sigma$.

1. $\hat{\delta}(q, \varepsilon) = \varepsilon\mathrm{H}(q)$
2. $\hat{\delta}(q, wa) = \bigcup\limits_{r \in \hat{\delta}(q,w)} \bigcup\limits_{p \in \delta(r,a)} \varepsilon\mathrm{H}(p)$



Vom ε-NEA zum NEA

- ε-NEA $(Q, \Sigma, \delta, q_0, F)$ $\longrightarrow$ NEA $(Q', \Sigma', \delta' q_0', F')$
- *Intuitiv:*



$$\delta'(q_2, 0) = \{q_4, q_6, f, q_1, q_2, q_3\} = \hat{\delta}(q_2, 0)$$

$$\delta'(q_0, 0) = \{q_4, q_6, f, q_1, q_2, q_3\} = \hat{\delta}(q_0, 0)$$

Außerdem: q_0 muss akzeptierend sein

Vom ε-NEA zum NEA

Satz 4.1: Zu jedem ε-NEA kann ein äquivalenter NEA ohne ε-Übergänge konstruiert werden.

Beweis: ε-NEA $\mathrm{A} = (Q, \Sigma, \delta, q_0, F) \longrightarrow$ NEA $\mathrm{A}' = (Q', \Sigma', \delta' q_0', F')$

mit $Q' = Q, \Sigma' = \Sigma, q_0' = q_0$,

$\delta'(q, a) = \hat{\delta}(q, a)$ für alle $q \in Q$ und $a \in \Sigma$,

$$F' = \begin{cases} F \cup \{q_0\} & \text{falls } \varepsilon\mathrm{H}(q_0) \cap F \neq \emptyset \\ F & \text{sonst} \end{cases}$$

$\longrightarrow$ $\varepsilon \in L(A')$ gdw. $\varepsilon \in L(A)$

Vom ε-NEA zum NEA

- *zu zeigen*: Für alle $w \in \Sigma^+$ gilt $\hat{\delta}'(q_0, w) = \hat{\delta}(q_0, w)$.
- Induktion über $|w|$.
- **I.A.:** $|\boldsymbol{w}| = \mathbf{1}.$ $w = a, a \in \Sigma$

$$\hat{\delta}'(q_0, a) = \bigcup_{p \in \hat{\delta}'(q_0, \varepsilon)} \delta'(p, a) = \delta'(q_0, a) = \hat{\delta}(q_0, a)$$

- **I.S.:** $|\boldsymbol{w}| > \mathbf{1}.$ $w = va, v \in \Sigma^+, a \in \Sigma$

$$\hat{\delta}'(q_0, va) = \bigcup_{p \in \hat{\delta}'(q_0, v)} \delta'(p, a) \overset{\textbf{I.V.}}{=} \bigcup_{p \in \hat{\delta}(q_0, v)} \delta'(p, a) = \bigcup_{p \in \hat{\delta}(q_0, v)} \hat{\delta}(p, a)$$

$$= \hat{\delta}(q_0, va)$$

Vom ε-NEA zum NEA

$$F' = \begin{cases} F \cup \{q_0\} & \text{falls } \varepsilon\mathrm{H}(q_0) \cap F \neq \emptyset \\ F & \text{sonst} \end{cases}$$

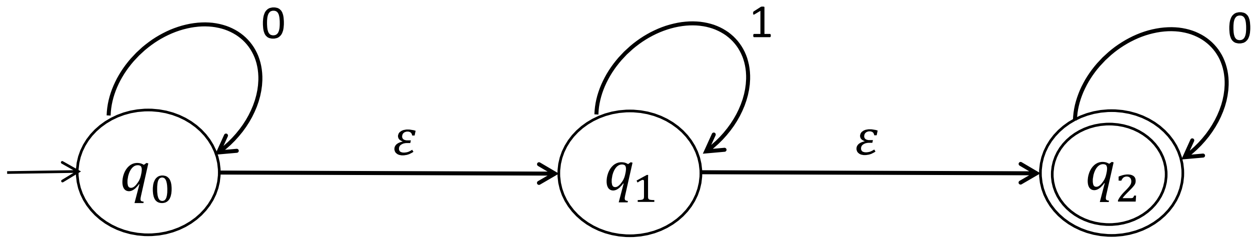
- *noch zu zeigen*: $\hat{\delta}'(q_0, w) \cap F' \neq \emptyset$ gdw. $\hat{\delta}(q_0, w) \cap F \neq \emptyset$

- $w = \varepsilon$: $\hat{\delta}'(q_0, \varepsilon) = \{q_0\}$. Es gilt $q_0 \in F'$ gdw. $\varepsilon\mathrm{H}(q_0) \cap F \neq \emptyset$
  gdw. $\hat{\delta}(q_0, \varepsilon) \cap F \neq \emptyset$

- $w = va$ für ein $a \in \Sigma$:
  - Falls $\hat{\delta}(q_0, w) \cap F \neq \emptyset$, dann $\hat{\delta}'(q_0, w) \cap F' \neq \emptyset$, da $F \subseteq F'$.
  - Falls $\hat{\delta}'(q_0, w) \cap (F' \setminus \{q_0\}) \neq \emptyset$, dann $\hat{\delta}(q_0, w) \cap F \neq \emptyset$.
  - Falls $q_0 \in \hat{\delta}'(q_0, w)$ und $q_0 \notin F$, dann existiert ein Zustand $f \in F$, der in der ε-Hülle von q_0 ist.

  ➢ Dann gilt $f \in \hat{\delta}(q_0, w)$. (*siehe Induktionsbeweis auf voriger Folie*) ∎

ε-NEA – Beispiel



$\varepsilon\mathrm{H}(q_0) = \{q_0, q_1, q_2\}$
$\varepsilon\mathrm{H}(q_1) = \{q_1, q_2\}$
$\varepsilon\mathrm{H}(q_2) = \{q_2\}$

$\hat{\delta}(q_0, \varepsilon) = \{q_0, q_1, q_2\}$ ⟶ $\boldsymbol{F' = F \cup \{q_0\} = \{q_0, q_2\}}$

$\hat{\delta}(q_0, 0)$: benötigen $\delta(q_0, 0) = \{q_0\}$, $\delta(q_1, 0) = \emptyset$, $\delta(q_2, 0) = \{q_2\}$

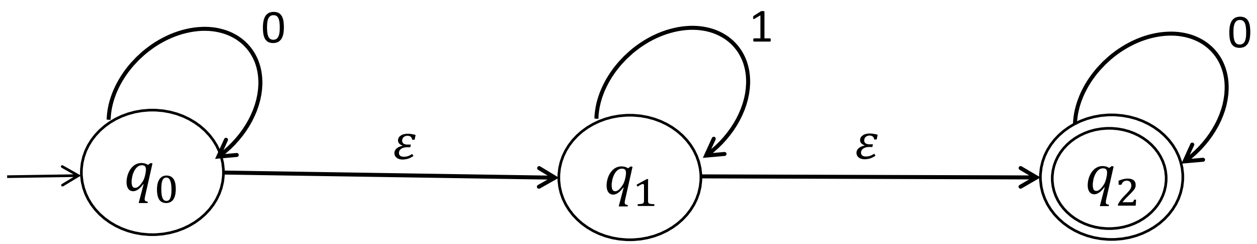
$\hat{\delta}(q_0, 0) = \varepsilon\mathrm{H}(q_0) \cup \emptyset \cup \varepsilon\mathrm{H}(q_2)$

$= \{q_0, q_1, q_2\}$ ⟶ $\boldsymbol{\delta'(q_0, 0) = \{q_0, q_1, q_2\}}$

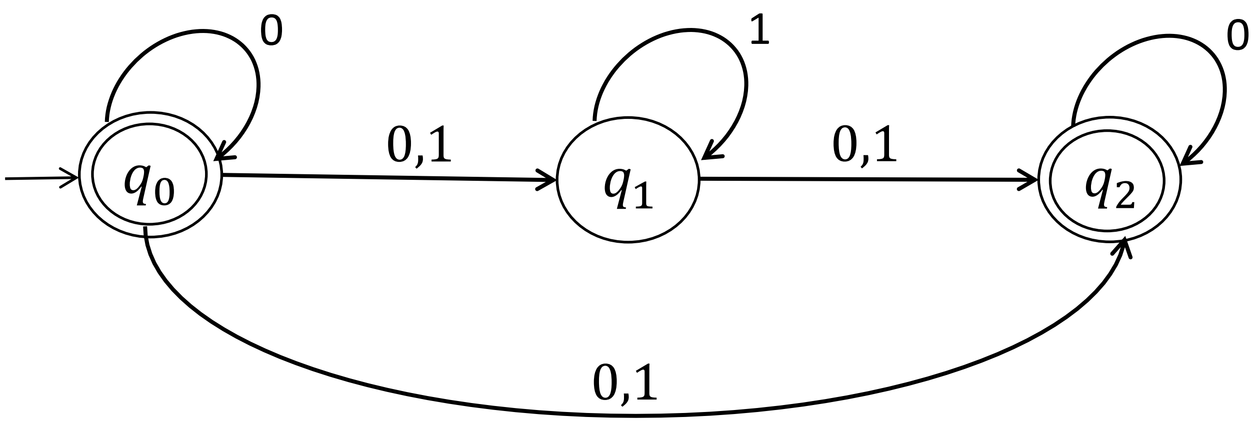
$\hat{\delta}(q_0, 1)$: benötigen $\delta(q_0, 1) = \emptyset$, $\delta(q_1, 1) = \{q_1\}$, $\delta(q_2, 1) = \emptyset$

$\hat{\delta}(q_0, 1) = \varepsilon\mathrm{H}(q_1) = \{q_1, q_2\}$ ⟶ $\boldsymbol{\delta'(q_0, 1) = \{q_1, q_2\}}$

Beseitigung von ε-Übergängen – Beispiel

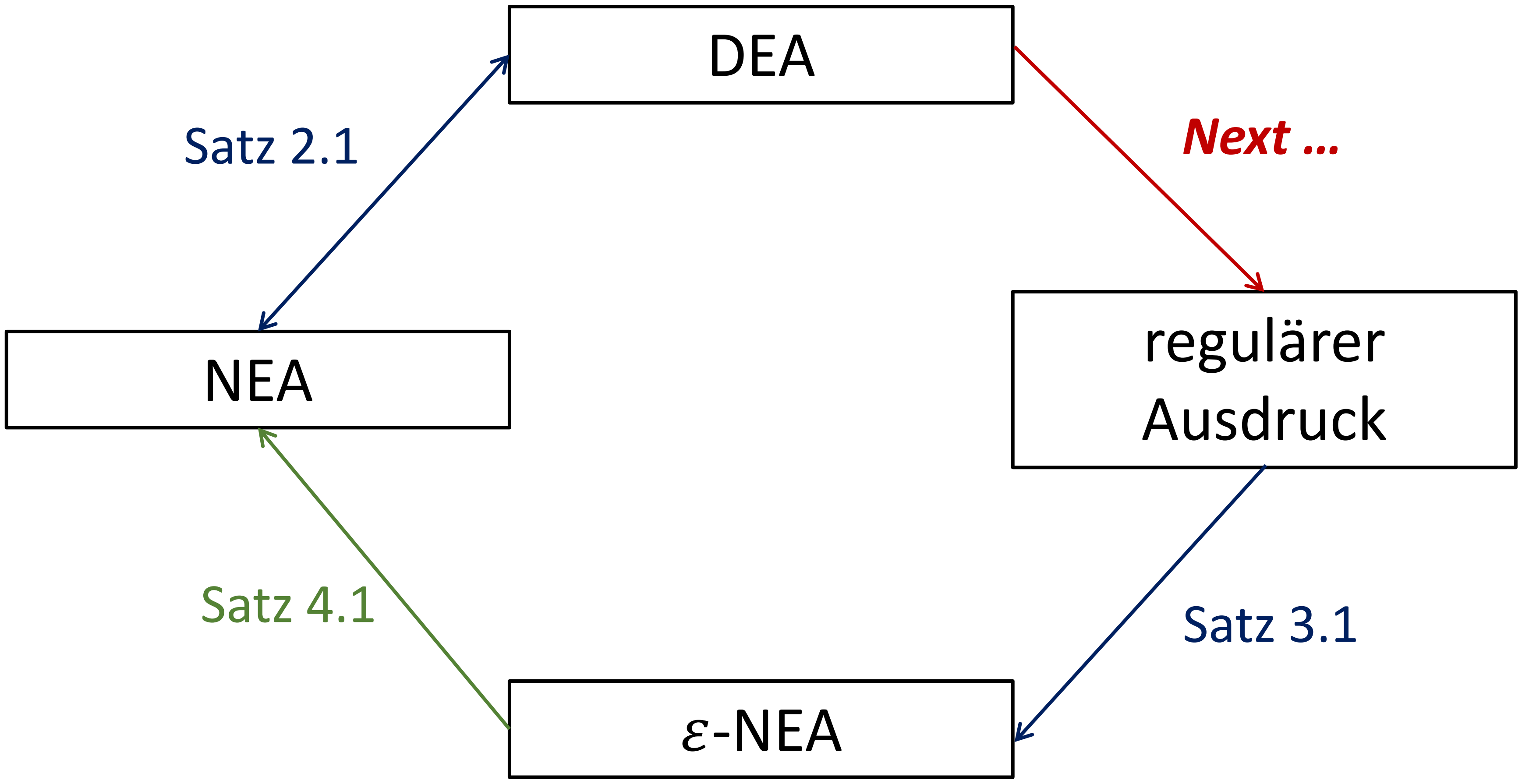


$$F' = F \cup \{q_0\} = \{q_0, q_2\}$$



	0	1
q_0	$\{q_0, q_1, q_2\}$	$\{q_1, q_2\}$
q_1	$\{q_2\}$	$\{q_1, q_2\}$
q_2	$\{q_2\}$	$\emptyset$

Äquivalenz der vier Mechanismen



Recap: Reguläre Ausdrücke – Definition

Definition: Sei Σ ein Alphabet.
Reguläre Ausdrücke (RA) über Σ sind rekursiv wie folgt definiert:

1. $\emptyset, \varepsilon, a$ sind reguläre Ausdrücke für alle $a \in \Sigma$
2. Wenn r und s reguläre Ausdrücke über Σ sind, dann auch $(r + s)$, (rs) und r^*

Beispiele für $\Sigma = \{0,1\}$

`0` `1` `(0+1)` `(0+1)*` `(1(0+1))`

`(0+1)*(1(0+1))`

Recap: Sprachen regulärer Ausdrücke

- **Definition:** Sei Σ ein Alphabet.
  Die **formalen Sprachen regulärer Ausdrücke** über Σ sind rekursiv wie folgt definiert:

1. $L(\emptyset) = \emptyset, \qquad L(\varepsilon) = \{\varepsilon\}, \qquad L(a) = \{a\}$ (für alle $a \in \Sigma$)
2. Sind r und s reguläre Ausdrücke über Σ, dann gilt

$$L(r + s) = L(r) \cup L(s)$$

$$L(rs) = L(r) \cdot L(s)$$

$$L(r^*) = L(r)^*$$

Von DEA zu regulären Ausdrücken

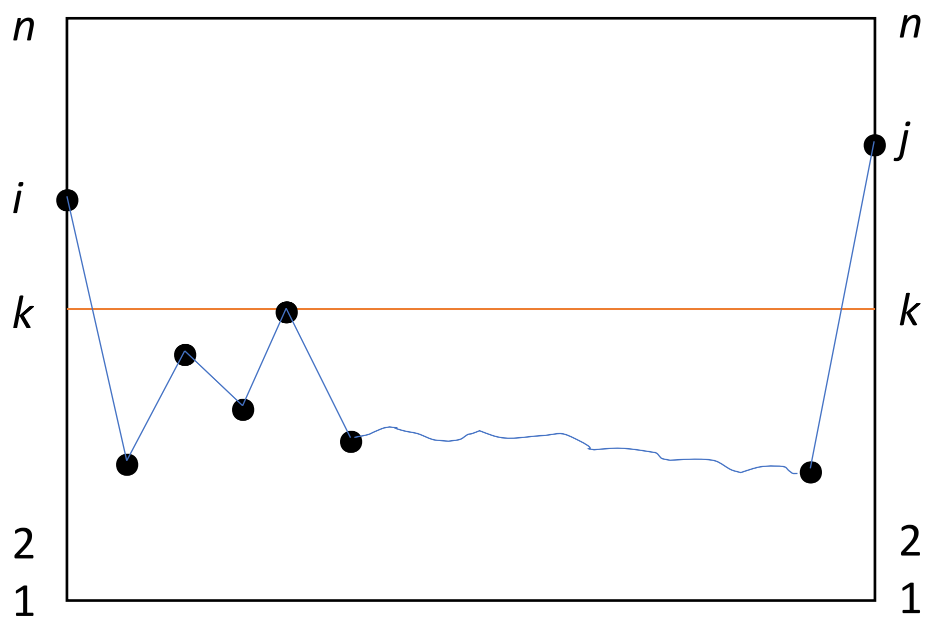
Satz 4.2: Zu jedem DEA A kann ein regulärer Ausdruck r_A mit $L(r_A) = L(A)$ konstruiert werden.

Beweis: Sei $A = (Q, \Sigma, \delta, q_1, F)$ ein DEA mit $Q = \{q_1, q_2, \dots, q_n\}$.

$\boldsymbol{R_{ij}^k}$ sei die Menge aller Eingabewörter, die von q_i nach q_j führen und dabei nur durch Zustände gehen, deren Index nicht größer als k ist.

R_{ij}^k ist die Menge aller $w \in \Sigma^*$, für die gilt:

1. $\hat{\delta}(q_i, w) = q_j$,
2. wenn u ein Präfix von w ist und $\hat{\delta}(q_i, u) = q_\ell$, dann ist $\ell \leq k$.



Von DEA zu regulären Ausdrücken

- Rekursive Definition von R_{ij}^k:

$$R_{ij}^k = R_{ik}^{k-1}\left(R_{kk}^{k-1}\right)^* R_{kj}^{k-1} \cup R_{ij}^{k-1}$$

$$R_{ij}^0 = \begin{cases} \{\, a \mid \delta(q_i, a) = q_j \} & \text{falls } i \neq j \\ \{\, a \mid \delta(q_i, a) = q_j \} \cup \{\varepsilon\} & \text{falls } i = j \end{cases}$$

- Für alle $1 \leq i, j, k \leq n$ existiert ein regulärer Ausdruck r_{ij}^k mit $L\left(r_{ij}^k\right) = R_{ij}^k$ **(VI über *k*)**:

  IA: Falls $R_{ij}^0 = \emptyset$: $r_{ij}^0 = \emptyset$. Falls $R_{ij}^0 = \{a_1, a_2, \dots, a_p\}$: $r_{ij}^0 = a_1 + a_2 + \cdots + a_p$.

  IS: Sei $k \geq 1$. Nach IV gibt es Ausdrücke r_{ij}^{k-1} für alle R_{ij}^{k-1}.

  Dann ist $r_{ij}^k = \left(r_{ik}^{k-1}\right)\left(r_{kk}^{k-1}\right)^*\left(r_{kj}^{k-1}\right) + r_{ij}^{k-1}$ Ausdruck für R_{ij}^k.

Von DEA zu regulären Ausdrücken

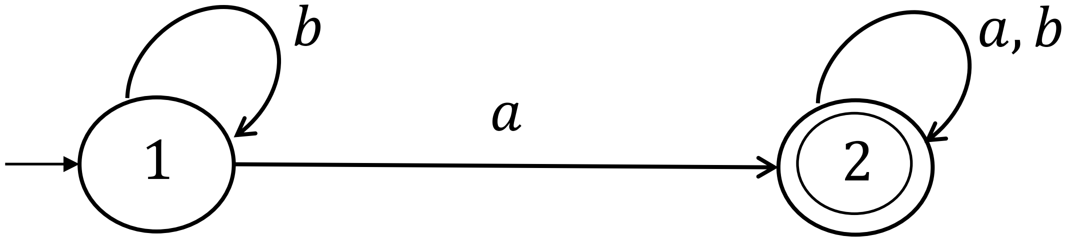
- Die Sprache des DEA A ist

$$L(A) = \bigcup_{q_j \in F} R_{1j}^n$$

- Falls $F = \{q_{j_1}, q_{j_2}, \dots, q_{j_m}\}$, dann gilt

$$L(A) = L(r_{1j_1}^n + r_{1j_2}^n + \cdots r_{1j_m}^n) \qquad \blacksquare$$

Beispiel



$\longrightarrow$ *intuitiv:* $b^*a(a+b)^*$

$$R_{ij}^0 = \begin{cases} \{\, a \mid \delta(q_i, a) = q_j \} & \text{falls } i \neq j \\ \{\, a \mid \delta(q_i, a) = q_j \} \cup \{\varepsilon\} & \text{falls } i = j \end{cases}$$

$$r_{11}^0 = b + \varepsilon \qquad r_{21}^0 = \emptyset$$
$$r_{12}^0 = a \qquad r_{22}^0 = a + b + \varepsilon$$

$$R_{ij}^k = R_{ik}^{k-1}\left(R_{kk}^{k-1}\right)^* R_{kj}^{k-1} \cup R_{ij}^{k-1}$$

$$r_{12}^2 = r_{12}^1 (r_{22}^1)^* r_{22}^1 + r_{12}^1$$

Beispiel

$$R_{ij}^{k} = R_{ik}^{k-1}\left(R_{kk}^{k-1}\right)^{*}R_{kj}^{k-1} \cup R_{ij}^{k-1}$$

$$\boldsymbol{r_{12}^{2} = r_{12}^{1}\left(r_{22}^{1}\right)^{*}r_{22}^{1} + r_{12}^{1}}$$

$$\boldsymbol{= b^{*}a(a+b+\varepsilon)^{*}(a+b+\varepsilon) + b^{*}a} \;\cong\; \boldsymbol{b^{*}a(a+b)^{*} + b^{*}a} \;\cong\; \boldsymbol{b^{*}a(a+b)^{*}}$$

$r_{11}^{0} = b + \varepsilon \qquad r_{21}^{0} = \emptyset$

$r_{12}^{0} = a \qquad r_{22}^{0} = a + b + \varepsilon$

$$r_{12}^{1} = r_{11}^{0}\left(r_{11}^{0}\right)^{*}r_{12}^{0} + r_{12}^{0}$$
$$= (b+\varepsilon)(b+\varepsilon)^{*}a + a \;\cong\; b^{*}a + a \;\cong\; b^{*}a$$

$$r_{22}^{1} = r_{21}^{0}\left(r_{11}^{0}\right)^{*}r_{12}^{0} + r_{22}^{0}$$
$$= \emptyset(b+\varepsilon)^{*}a + (a+b+\varepsilon) \;\cong\; \emptyset + a + b + \varepsilon \;\cong\; a + b + \varepsilon$$

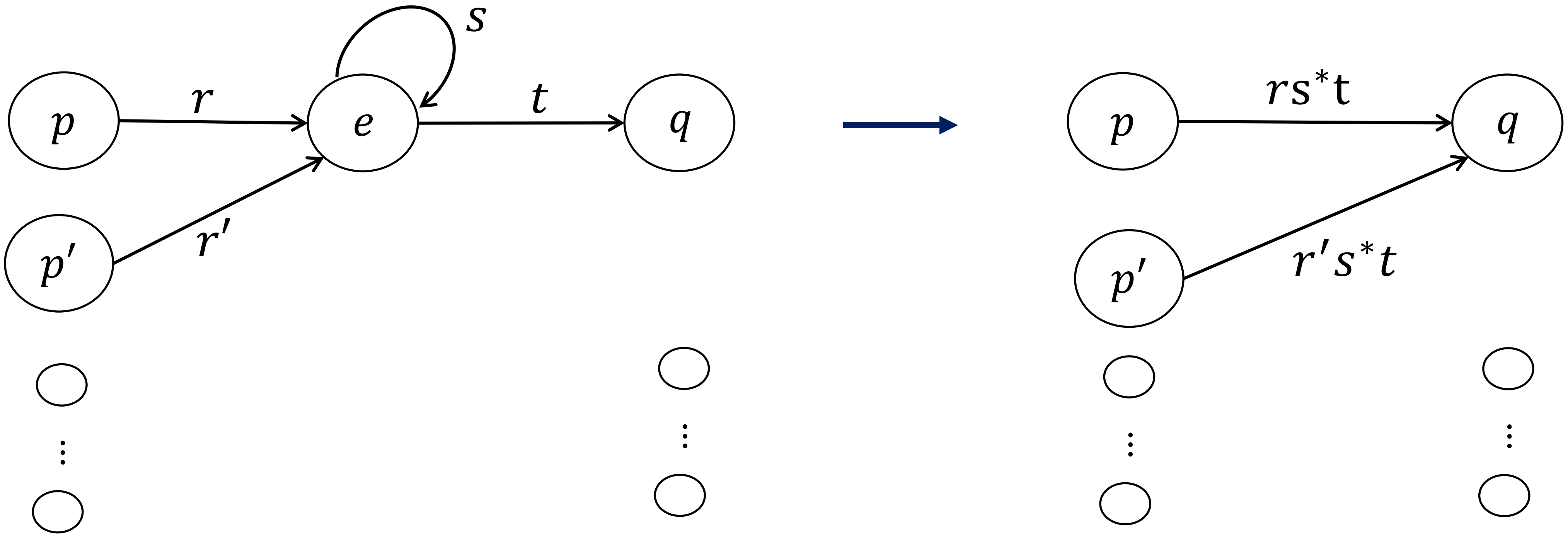
Ausblick: Zustandseliminierung

- ein intuitiveres Verfahren, das oft kleinere Ausdrücke liefert
- nutzen **RA-Automaten**:
  erlauben reguläre Ausdrücke als Kantenmarkierungen in Automaten
- Jeder (D)EA kann als spezieller RA-Automat aufgefasst werden!
  - Symbol an Kante als regulärer Ausdruck
  - mehrere Symbole (z.B. a, b) als regulärer Ausdruck $(a + b)$
  - sogar ε an Kante kann als regulärer Ausdruck aufgefasst werden
- Im Folgenden gehen wir von einem beliebigen RA-Automaten aus.

Ausblick: Zustandseliminierung

Für jeden akzeptierenden Zustand q_f:

Für jeden Zustand e, der vom Startzustand und q_f verschieden ist:



Ausblick: Zustandseliminierung

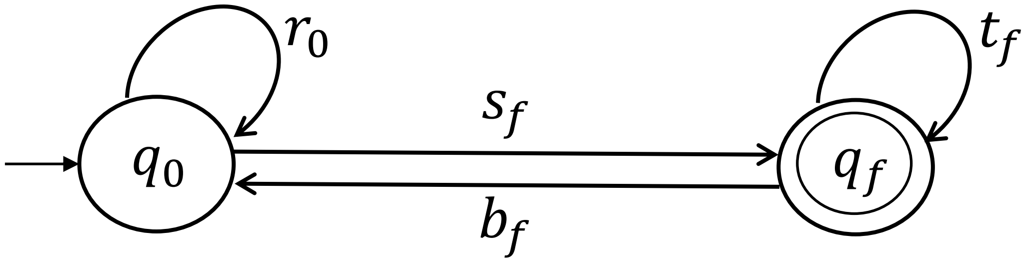
Für jeden akzeptierenden Zustand q_f:

Für jeden Zustand e, der vom Startzustand und q_f verschieden ist:

Eliminiere e

bis Automaten der Form

entstanden sind.

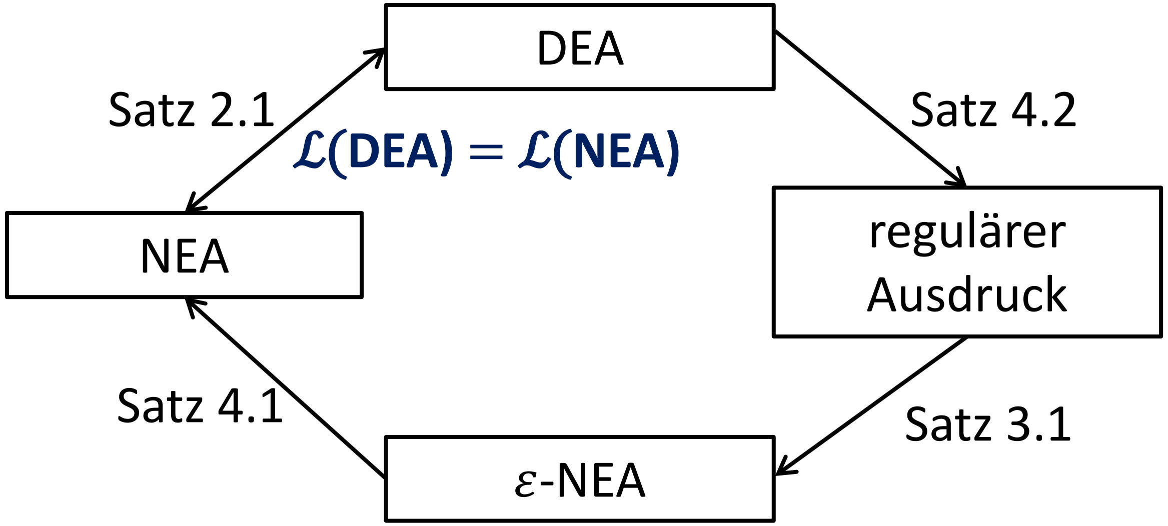


Für $F = \{q_{f_1}, q_{f_2}, \dots, q_{f_m}\}$ bilde den Ausdruck $re = a_{f_1} + a_{f_2} + \cdots + a_{f_m}$ mit

$$a_{f_i} = r_0^* s_{f_i} (t_{f_j} + b_{f_j} (r_0)^* s_{f_j})^*$$

Dann gilt $L(re) = L(A)$. *(ohne formalen Beweis)*

Reguläre Sprachen



Folgerung 4.3:

$\mathcal{L}$(DEA) = $\mathcal{L}$(NEA) = $\mathcal{L}$(ε-NEA) = $\mathcal{L}$(REG)

$\mathcal{L}$(ε-NEA) = $\{\, L \mid L = L(A)$ für einen ε-NEA $A \,\}$

$\mathcal{L}$(REG) = $\{\, L \mid L = L(r)$ für einen regulären Ausdruck $r \,\}$