Karlsruher Institut für Technologie

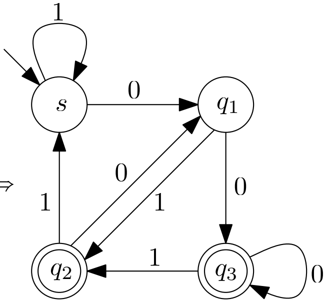
Theoretische Grundlagen der Informatik
Vorlesung am 19. Oktober 2017

INSTITUT FÜR THEORETISCHE INFORMATIK

Endliche Automaten und Reguläre Sprachen

- Beispiel: Sprache aller Wörter über $\{0, 1\}$, deren vorletztes Zeichen eine 0 ist:

$$L := (0 \cup 1)^* 0 (0 \cup 1) \quad \Leftrightarrow$$



- Gibt es zu jeder regulären Sprache einen endlichen Automaten, der diese erkennt?
- Falls ja: Wie kann man diesen konstruieren?
- Welche Sprachen werden durch endliche Automaten erkannt?

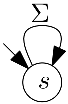
Satz:
Jede reguläre Sprache wird von einem (deterministischen) endlichen Automaten (DEA) akzeptiert.

Erinnerung: Eine Sprache $L \subseteq \Sigma^*$ heißt *regulär*, wenn für sie einer der folgenden Punkte gilt: (induktive Definition)

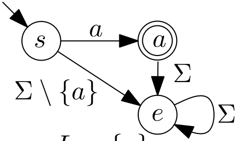
- Verankerung:
  - $L = \{a\}$ mit $a \in \Sigma$ oder
  - $L = \{\varepsilon\}$ oder
  - $L = \emptyset$
- Induktion: Seien L_1, L_2 reguläre Sprachen
  - $L = L_1 \cdot L_2$ oder
  - $L = L_1 \cup L_2$ oder
  - $L = L_1^*$

Satz:
Jede reguläre Sprache wird von einem (deterministischen) endlichen Automaten (DEA) akzeptiert.

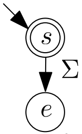
- $L := L(\alpha)$ reguläre Sprache über Σ, die durch α beschreibbar ist
- Induktion über $n =$Anzahl der $\cup$, $\cdot$ und *-Zeichen in α
- **Induktionsanfang** $n = 0$:



$L = \emptyset$



$L = \{a\}$

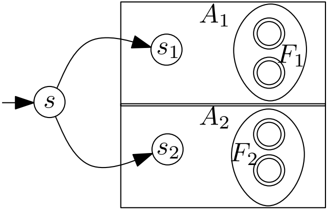


$L = \{\varepsilon\}$

- **Induktionsannahme**: Für alle L, die durch reguläre Ausdrücke mit weniger als n Operationen beschreibbar sind, existiert akzeptierender DEA

Satz:
Jede reguläre Sprache wird von einem (deterministischen) endlichen Automaten (DEA) akzeptiert.

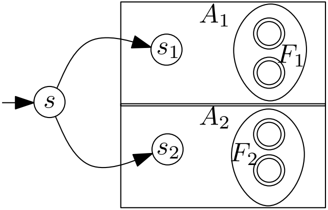
- **Induktionsschluss**:
  - $n > 0 \Rightarrow L = L_1 \cup L_2$ oder $L = L_1 \cdot L_2$ oder $L = L_1^*$
  - **Fall 1:** $L = L_1 \cup L_2$
  - Sei A_1 DEA der L_1 akzeptiert und A_2 DEA der L_2 akzeptiert
  - Idee: „Baue“ aus A_1 und A_2 Automaten der L akzeptiert
  - 1. Versuch:



- Brauche Übergang ohne Lesen eines Zeichens mit Wahlmöglichkeit
- DEA's erlauben das nicht
- Rest des Beweises: später!

Satz:
Jede reguläre Sprache wird von einem (deterministischen) endlichen Automaten (DEA) akzeptiert.

- **Induktionsschluss**:
  - $n > 0 \Rightarrow L = L_1 \cup L_2$ oder $L = L_1 \cdot L_2$ oder $L = L_1^*$
  - **Fall 1:** $L = L_1 \cup L_2$
  - Sei A_1 DEA der L_1 akzeptiert und A_2 DEA der L_2 akzeptiert
  - Idee: „Baue" aus A_1 und A_2 Automaten der L akzeptiert
  - 1. Versuch:



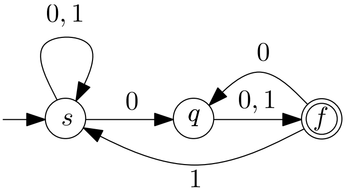
- Brauche Übergang ohne Lesen eines Zeichens mit Wahlmöglichkeit
- DEA's erlauben das nicht
- Rest des Beweises: später!

Nichtdeterministische endliche Automaten

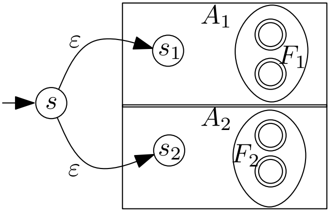
- Ein **nichtdeterministischer endlicher Automat** (NEA) besteht aus:
  - Q, einer endlichen Zustandsmenge;
  - Σ, einem endlichen Alphabet;
  - δ, einer Übergangsfunktion $\delta\colon Q \times (\Sigma \cup \{\varepsilon\}) \to 2^Q$
  - s, einem Startzustand;
  - F, einer Menge von Endzuständen;
- Interpretation von δ:
  - Bei Abarbeitung eines Symbols a von Zustand q kann der Automat sich **nichtdeterministisch „aussuchen“** in welchen Zustand aus der Teilmenge $\delta(q, a)$ er übergeht
  - Auch möglich: Springe spontan ohne Lesen eines Zeichens aus Σ in neuen Zustand aus Teilmenge $\delta(q, \varepsilon)$ (ε-Übergang)
- Ein nichtdeterministischer endlicher Automat **akzeptiert** eine Wort $w \in \Sigma^*$, wenn es eine Folge von Übergängen gibt, so dass er bei Eingabe von w in einen Endzustand gelangt, d.h. bei Eingabe von w ein Endzustand *erreichbar* ist.

Beispiele für NEA's

Beispiel 1:

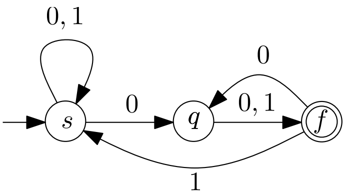


Beispiel 2:



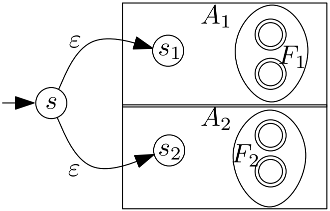
Beispiele für NEA's

Beispiel 1:



- $\delta(s, 0) = \{s, q\}$
- $w = 10$ wird nicht akzeptiert
- $w = 01$ wird akzeptiert
- Sprache aller Wörter über $\{0, 1\}$, deren vorletztes Zeichen 0 ist

Beispiel 2:



Äquivalenz von NEA's und DEA's

Definition:
Zwei endliche Automaten, die dieselbe Sprache akzeptieren, heißen *äquivalent*.

Satz:
Zu jedem nichtdeterministischen endlichen Automaten gibt es einen äquivalenten deterministischen endlichen Automaten.

Beweis durch Potenzmengenkonstruktion

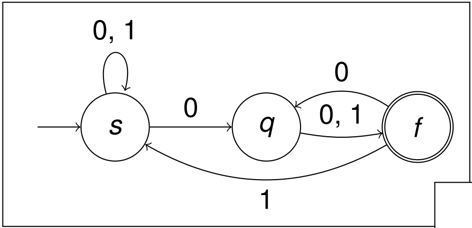
Äquivalenz von NEA's und DEA's

Definition:
Zwei endliche Automaten, die dieselbe Sprache akzeptieren, heißen *äquivalent*.

Satz:
Zu jedem nichtdeterministischen endlichen Automaten gibt es einen äquivalenten deterministischen endlichen Automaten.

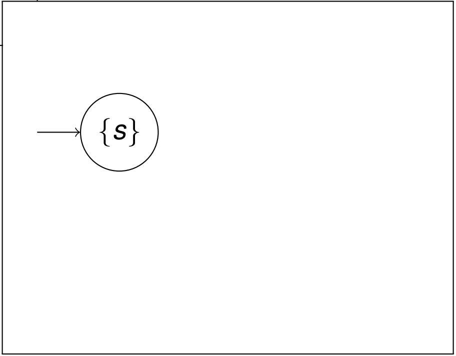
Beweis durch Potenzmengenkonstruktion

Beispiel Potenzmengenkonstruktion

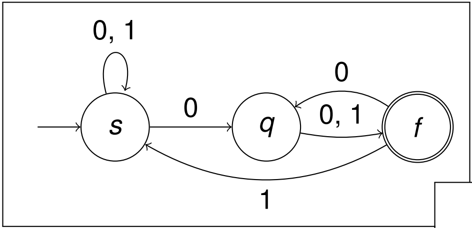


$\leftarrow$ NEA

DEA $\rightarrow$

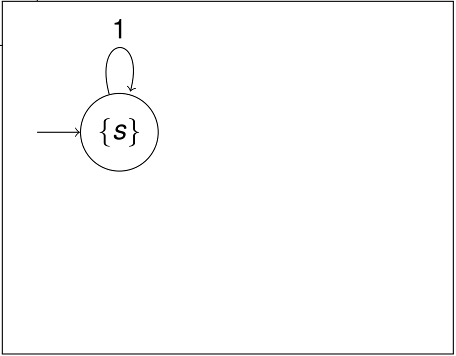


Beispiel Potenzmengenkonstruktion

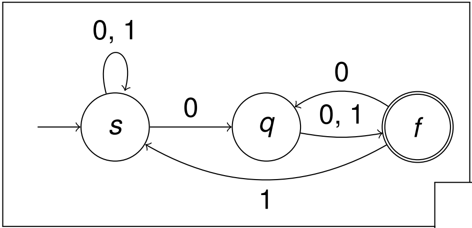


← NEA

DEA →

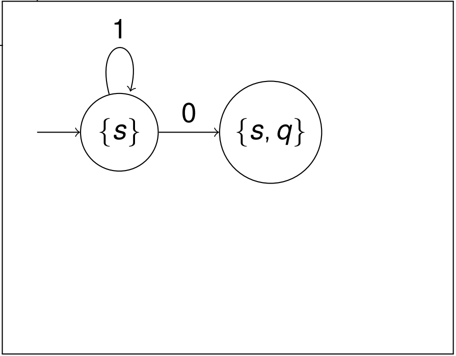


Beispiel Potenzmengenkonstruktion

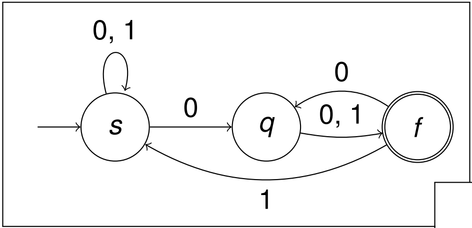


← NEA

DEA →

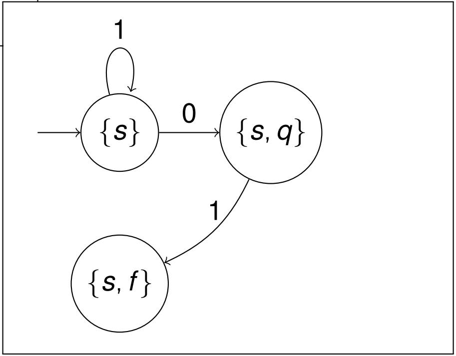


Beispiel Potenzmengenkonstruktion

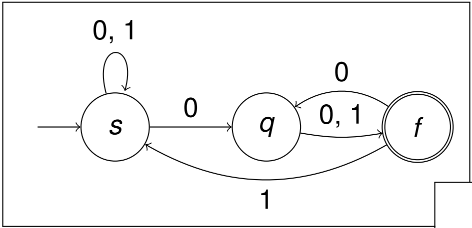


← NEA

DEA →

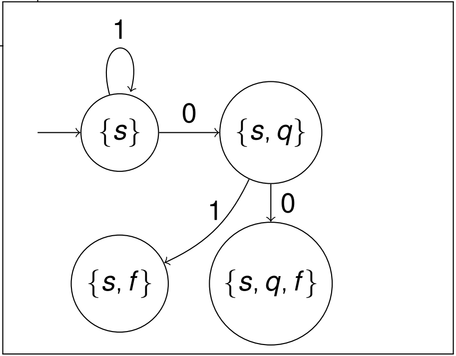


Beispiel Potenzmengenkonstruktion

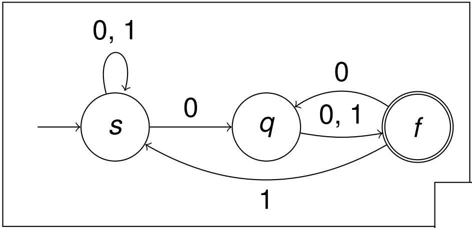


← NEA

DEA →

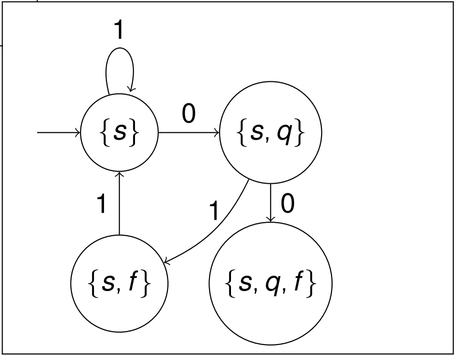


Beispiel Potenzmengenkonstruktion

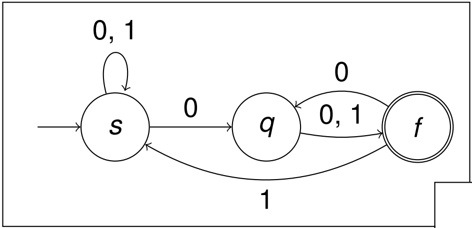


← NEA

DEA →

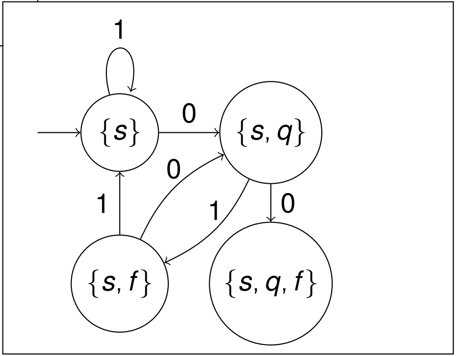


Beispiel Potenzmengenkonstruktion

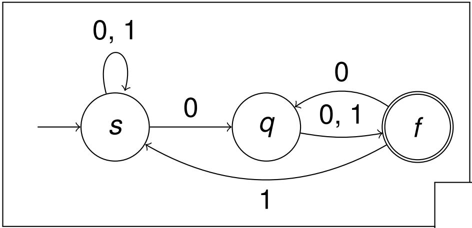


← NEA

DEA →

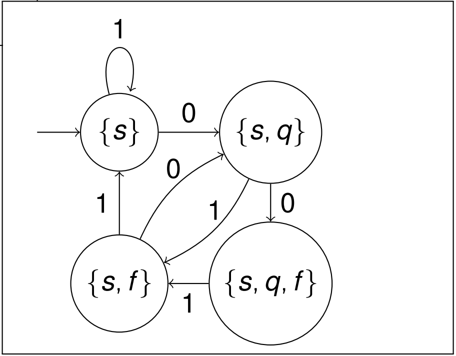


Beispiel Potenzmengenkonstruktion

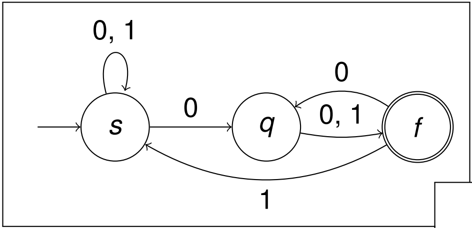


$\leftarrow$ NEA

DEA $\rightarrow$

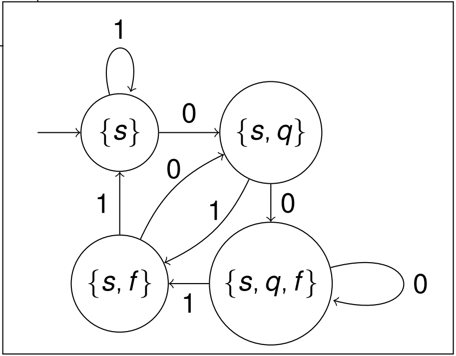


Beispiel Potenzmengenkonstruktion

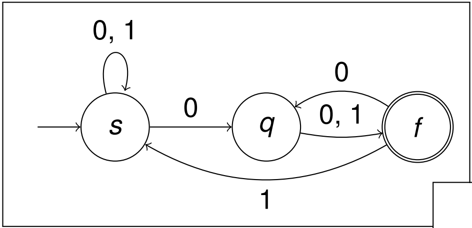


$\leftarrow$ NEA

DEA $\rightarrow$

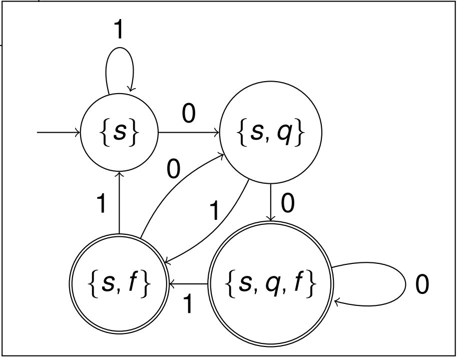


Beispiel Potenzmengenkonstruktion



$\leftarrow$ NEA

DEA $\rightarrow$



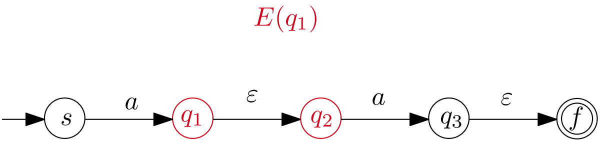
Erweiterung von δ

Definition:
Für einen Zustand $q \in Q$ ist der ε–**Abschluss** $E(q)$ wie folgt definiert:

$$E(q) := \{p \in Q \mid p \text{ ist von } q \text{ durch } \varepsilon\text{–Übergänge erreichbar}\}$$

Beachte, dass gilt:

- $E(q) \subseteq Q, \quad E(q) \in 2^Q$
- $q \in E(q)$, d.h. die Folge von ε-Übergängen darf auch leer sein



Erweiterung $\bar{\delta}$ von δ

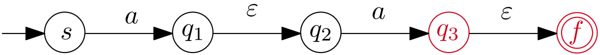
Idee: Berücksichtige ε-Übergänge bei der Übergangsfunktion

- Beim Lesen eines einzelnen Zeichens

$$\bar{\delta}\colon Q \times (\Sigma \cup \{\varepsilon\}) \to 2^Q$$

$$\bar{\delta}(q,a) = \begin{cases} E(q) & \text{falls } a = \varepsilon \\ \bigcup\limits_{p \in E(q)} \left(\bigcup\limits_{r \in \delta(p,a)} E(r) \right) & \text{für } a \in \Sigma \end{cases}$$

$\bar{\delta}(q_1, a)$



Erweiterung $\bar{\delta}$ von δ

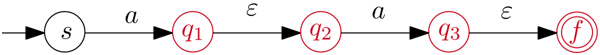
Idee: Berücksichtige ε-Übergänge bei der Übergangsfunktion

- Erweiterung auf Mengen von Zuständen :

$$\bar{\delta}\colon 2^Q \times (\Sigma \cup \{\varepsilon\}) \to 2^Q$$

$$\bar{\delta}(P, a) = \begin{cases} \bigcup\limits_{p \in P} E(p) & \text{falls } a = \varepsilon \\ \bigcup\limits_{p \in P} \bar{\delta}(p, a) & \text{für } a \in \Sigma \end{cases}$$

$\bar{\delta}(\{s, q_1\}, a)$



Erweiterung $\bar{\delta}$ von δ

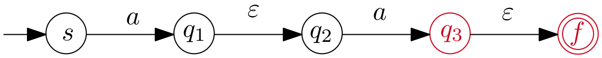
Idee: Berücksichtige ε-Übergänge bei der Übergangsfunktion

- Induktive Erweiterung auf Wörter:

$$\bar{\delta}\colon Q \times \Sigma^* \to 2^Q$$

$$\bar{\delta}(q,w) = \begin{cases} E(q) & \text{falls } w = \varepsilon \\ \bar{\delta}(q,w) & \text{falls } w = a \in \Sigma \\ \bigcup\limits_{p \in \bar{\delta}(q,v)} \bar{\delta}(p,a) & \text{falls } w = va,\ a \in \Sigma,\ |v| > 0 \end{cases}$$

$$\bar{\delta}(s, w = aa)$$



- Es gilt: $w \in L(\mathcal{A}) \Leftrightarrow \bar{\delta}(s,w) \cap F \neq \emptyset$!

Erweiterung $\bar{\delta}$ von δ

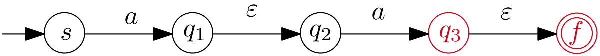
Idee: Berücksichtige ε-Übergänge bei der Übergangsfunktion

- Analog für Mengen von Zuständen:

$$\bar{\delta} \colon 2^Q \times \Sigma^* \to 2^Q$$

$$\bar{\delta}(P, w) = \bigcup_{p \in P} \bar{\delta}(p, w)$$

$$\bar{\delta}(\{s, q_1\}, w = aa)$$



Äquivalenz von NEA's und DEA's

Definition:
Zwei endliche Automaten, die dieselbe Sprache akzeptieren, heißen *äquivalent*.

Satz:
Zu jedem nichtdeterministischen endlichen Automaten gibt es einen äquivalenten deterministischen endlichen Automaten.

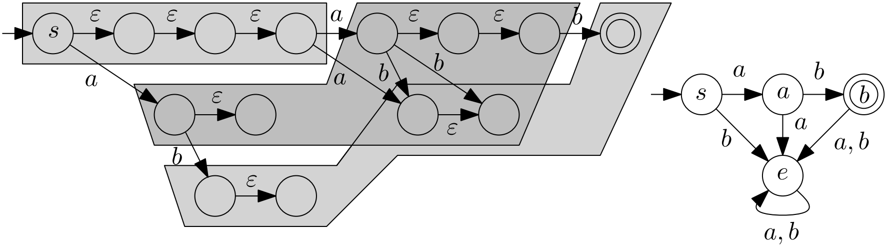
Satz:
Zu jedem nichtdeterministischen endlichen Automaten gibt es einen äquivalenten deterministischen endlichen Automaten.

Beweisidee:

- Sei $\mathcal{A}$ NEA
- Falls $\mathcal{A}$ w akzeptiert, so gibt es eine Abarbeitung, die in einem Endzustand endet
- Falls ein Zustand q bei der Eingabe von w erreichbar ist, so sind das auch alle Zustände in $E(q)$
- „Simuliere“ $\mathcal{A}$ durch einen DEA
- Zustände des DEA sind Mengen von Zuständen von $\mathcal{A}$
- Endzustände des DEA sind alle Mengen, die Endzustände von $\mathcal{A}$ enthalten

Satz:
Zu jedem nichtdeterministischen endlichen Automaten gibt es einen äquivalenten deterministischen endlichen Automaten.

Beispiel: Mögliche Abarbeitungen von $w = ab$ in einem NEA mit $ab \in L$, $a \notin L$ und äquivalenter DEA



Satz:
Zu jedem nichtdeterministischen endlichen Automaten gibt es einen äquivalenten deterministischen endlichen Automaten.

Potenzmengenkonstruktion: Gegeben sei ein NEA $\mathcal{A} := (Q, \Sigma, \delta, s, F)$. Wir konstruieren daraus einen DEA $\widetilde{\mathcal{A}} := (\widetilde{Q}, \Sigma, \widetilde{\delta}, \widetilde{s}, \widetilde{F})$:

- $\widetilde{Q} = 2^Q$, d.h. die Zustände des DEA sind Mengen von Zuständen des NEA.
- $\widetilde{\delta}\colon \widetilde{Q} \times \Sigma \to \widetilde{Q}$ mit $\widetilde{\delta}(\widetilde{q}, a) = \bar{\delta}(\widetilde{q}, a)$ für $a \in \Sigma$. Es ist also $\widetilde{q} \subseteq Q$ und jeder Zustand wird mit seinem ε–Abschluss im NEA identifiziert.
- $\widetilde{s} := E(s)$
- $\widetilde{F} := \{\widetilde{q} \in \widetilde{Q} \mid \widetilde{q} \cap F \neq \emptyset\}$

Satz:
Zu jedem nichtdeterministischen endlichen Automaten gibt es einen äquivalenten deterministischen endlichen Automaten.

- Zeige per Induktion über $n = |w|$, dass für alle $w \in \Sigma^*$ gilt:

$$\widetilde{\delta}(\widetilde{s}, w) = \bar{\delta}(s, w)$$

- Dann gilt:

$$\begin{aligned} w \in L(\widetilde{\mathcal{A}}) \Leftrightarrow \widetilde{\delta}(\widetilde{s}, w) \in \widetilde{F} \Leftrightarrow \widetilde{\delta}(\widetilde{s}, w) \cap F \neq \emptyset \\ \Leftrightarrow \bar{\delta}(s, w) \cap F \neq \emptyset \Leftrightarrow w \in L(\mathcal{A}) \end{aligned}$$

- **Induktionsanfang** $n = 0$, d.h. $w = \varepsilon$:

$$\widetilde{\delta}(\widetilde{s}, \varepsilon) = \bar{\delta}(E(s), \varepsilon) = \cup_{p \in E(s)} E(p) = E(s) = \bar{\delta}(s, \varepsilon)$$

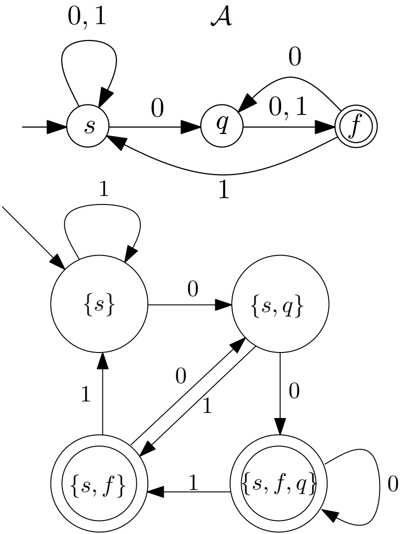
Satz:
Zu jedem nichtdeterministischen endlichen Automaten gibt es einen äquivalenten deterministischen endlichen Automaten.

- **Induktionsannahme**: Für alle Wörter w' mit $|w'| \leq n$ gilt:
  $\widetilde{\delta}(\widetilde{s}, w') = \bar{\delta}(s, w')$
- **Induktionsschluss** $|w| = n + 1$:
  Sei $w = w'a$ mit $|w'| = n$ und $a \in \Sigma$. Dann gilt:

$$
\begin{aligned}
\widetilde{\delta}(\widetilde{s}, w) &= \widetilde{\delta}(\widetilde{\delta}(\widetilde{s}, w'), a) \\
&\underset{(\mathrm{IV})}{=} \widetilde{\delta}(\bar{\delta}(s, w'), a) \\
&\underset{\mathrm{Def}\ \widetilde{\delta}}{=} \bar{\delta}(\bar{\delta}(s, w'), a) = \bigcup_{p \in \bar{\delta}(s, w')} \bar{\delta}(p, a) = \bar{\delta}(s, w)
\end{aligned}
$$

Beispiel Potenzmengenkonstruktion:

- $L = (0 \cup 1)^* 0 (0 \cup 1)$
- NEA $\mathcal{A}$ akzeptiert L
- Konstruiere DEA für L:
  - $\widetilde{s} = E(s) = \{s\}$
  - $\widetilde{\delta}(\{s\}, 0) = \{s, q\}$
  - $\widetilde{\delta}(\{s\}, 1) = \{s\}$
  - $\widetilde{\delta}(\{s, q\}, 0) = \{s, q, f\}$
  - $\widetilde{\delta}(\{s, q\}, 1) = \{s, f\}$
  - $\widetilde{\delta}(\{s, q, f\}, 0) = \{s, q, f\}$
  - $\widetilde{\delta}(\{s, q, f\}, 1) = \{s, f\}$
  - $\widetilde{\delta}(\{s, f\}, 0) = \{s, q\}$
  - $\widetilde{\delta}(\{s, f\}, 1) = \{s\}$
  - Alle anderen Zustände können gestrichen werden



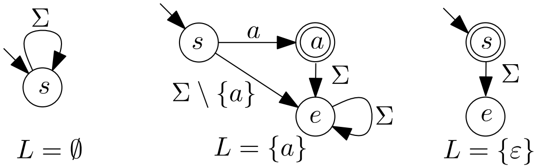
Satz:
Jede reguläre Sprache wird von einem (deterministischen) endlichen Automaten (DEA) akzeptiert.

Erinnerung: Eine Sprache $L \subseteq \Sigma^*$ heißt *regulär*, wenn für sie einer der folgenden Punkte gilt: (induktive Definition)

- Verankerung:
  - $L = \{a\}$ mit $a \in \Sigma$ oder
  - $L = \emptyset$
  - $L = \{\varepsilon\}$
- Induktion: Seien L_1, L_2 reguläre Sprachen
  - $L = L_1 \cdot L_2$ oder
  - $L = L_1 \cup L_2$ oder
  - $L = L_1^*$

Satz:
Jede reguläre Sprache wird von einem (deterministischen) endlichen Automaten (DEA) akzeptiert.

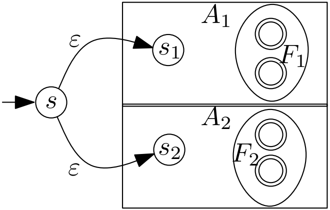
- $L := L(\alpha)$ reguläre Sprache über Σ, die durch α beschreibbar ist
- Induktion über $n =$Anzahl der $\cup$, $\cdot$ und *-Zeichen in α
- **Induktionsanfang** $n = 0$:



- **Induktionsannahme**: Für alle L, die durch reguläre Ausdrücke mit weniger als n Operationen beschreibbar sind, existiert akzeptierender DEA

Satz:
Jede reguläre Sprache wird von einem (deterministischen) endlichen Automaten (DEA) akzeptiert.

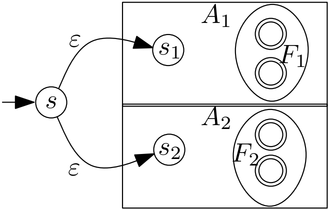
- **Induktionsschluss**:
  - $n > 0 \Rightarrow L = L_1 \cup L_2$ oder $L = L_1 \cdot L_2$ oder $L = L_1^*$
  - **Fall 1:** $L = L_1 \cup L_2$
  - Sei A_1 DEA, der L_1 akzeptiert und A_2 DEA, der L_2 akzeptiert
  - Idee: „Baue" aus A_1 und A_2 Automaten, der L akzeptiert



- NEA's können Übergang ohne Lesen eines Zeichens mit Wahlmöglichkeit
- NEA's können in DEA's umgebaut werden

Satz:
Jede reguläre Sprache wird von einem (deterministischen) endlichen Automaten (DEA) akzeptiert.

- **Induktionsschluss**:
  - $n > 0 \Rightarrow L = L_1 \cup L_2$ oder $L = L_1 \cdot L_2$ oder $L = L_1^*$
  - **Fall 1:** $L = L_1 \cup L_2$
  - Sei A_1 DEA, der L_1 akzeptiert und A_2 DEA, der L_2 akzeptiert
  - Idee: „Baue" aus A_1 und A_2 Automaten, der L akzeptiert



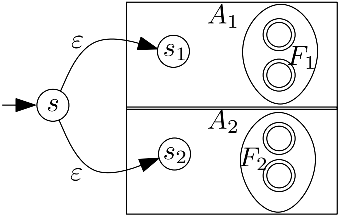
- NEA's können Übergang ohne Lesen eines Zeichens mit Wahlmöglichkeit
- NEA's können in DEA's umgebaut werden

Satz:
Jede reguläre Sprache wird von einem (deterministischen) endlichen Automaten (DEA) akzeptiert.

Beweis (Fortsetzung):

- **Fall 1:** $L = L_1 \cup L_2$
- NEA $\mathcal{A} := (Q, \Sigma, \delta, s, F)$, der $L_1 \cup L_2$ erkennt:
  - $Q := Q_1 \cup Q_2 \cup \{s\}$ $(s \notin Q_i)$
  - $F = F_1 \cup F_2$
  - δ :

$$\delta(q, a) := \begin{cases} \{\delta_i(q, a)\} & , q \in Q_i,\ a \in \Sigma \\ \emptyset & , q \in Q \setminus \{s\},\ a = \varepsilon \\ \{s_1, s_2\} & , q = s,\ a = \varepsilon \\ \emptyset & , q = s,\ a \neq \varepsilon \end{cases}$$

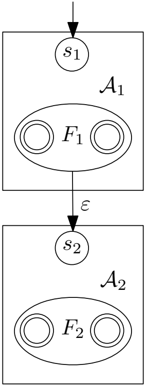


Satz:
Jede reguläre Sprache wird von einem (deterministischen) endlichen Automaten (DEA) akzeptiert.

Beweis (Fortsetzung):

- **Fall 2:** $L = L_1 \cdot L_2$
- NEA $\mathcal{A} := (Q, \Sigma, \delta, s, F)$, der $L_1 \cdot L_2$ erkennt:
  - $Q := Q_1 \cup Q_2$
  - $s := s_1$
  - $F := F_2$
  - δ:

$$\delta(q, a) := \begin{cases} \{\delta_i(q, a)\} & , q \in Q_i,\ a \in \Sigma \\ \emptyset & , q \in Q \setminus F_1,\ a = \varepsilon \\ \{s_2\} & , q \in F_1,\ a = \varepsilon \end{cases}$$



INSTITUT FÜR THEORETISCHE INFORMATIK

Satz:
Jede reguläre Sprache wird von einem (deterministischen) endlichen Automaten (DEA) akzeptiert.

Beweis (Fortsetzung):

- **Fall 3:** $L = L_1^*$
- NEA $\mathcal{A} := (Q, \Sigma, \delta, s, F)$, der L_1^* erkennt:
  - $Q := Q_1 \cup \{s, f\}$, $(s, f \notin Q_1)$
  - δ:

$$\delta(q, a) := \begin{cases} \{\delta_1(q, a)\} & , q \in Q_1,\ a \in \Sigma \\ \emptyset & , q \in Q_1 \setminus F_1,\ a = \varepsilon \\ \{f\} & , q \in F_1 \cup \{s\},\ a = \varepsilon \\ \emptyset & , q \in \{s, f\},\ a \neq \varepsilon \\ \{s_1\} & , q \in \{s, f\},\ a = \varepsilon \end{cases}$$