LMU
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Formale Sprachen und Komplexität
Theoretische Informatik für Medieninformatiker
Sommersemester 2023

Der Satz von Kleene

Prof. Dr. Jasmin Blanchette

Lehr- und Forschungseinheit für
Theoretische Informatik

Stand: 9. Mai 2023
Folien ursprünglich von PD Dr. David Sabel

Wiederholung: Reguläre Ausdrücke

Definition (Regulärer Ausdruck)

Sei Σ ein Alphabet. Ein regulärer Ausdruck über Σ ist induktiv definiert:

- $\emptyset$ ist ein regulärer Ausdruck
- ε ist ein regulärer Ausdruck
- a mit $a \in \Sigma$ ist ein regulärer Ausdruck
- Wenn α_1 und α_2 reguläre Ausdrücke sind, dann auch $\alpha_1\alpha_2$
- Wenn α_1 und α_2 reguläre Ausdrücke sind, dann auch $(\alpha_1|\alpha_2)$
- Wenn α ein regulärer Ausdruck ist, dann auch $(\alpha)^*$

Erzeugte Sprache

Die von einem regulären Ausdruck α erzeugte Sprache $L(\alpha)$ ist rekursiv definiert:

$$\begin{aligned}
L(\emptyset) &:= \emptyset \\
L(\varepsilon) &:= \{\varepsilon\} \\
L(a) &:= \{a\} \quad \text{für } a \in \Sigma \\
L(\alpha_1\alpha_2) &:= L(\alpha_1)L(\alpha_2) \\
&\;= \{uv \mid u \in L(\alpha_1), v \in L(\alpha_2)\} \\
L(\alpha_1|\alpha_2) &:= L(\alpha_1) \cup L(\alpha_2) \\
L((\alpha)^*) &:= L(\alpha)^*
\end{aligned}$$

Satz von Kleene

Theorem 4.7.4 (Satz von Kleene)

Reguläre Ausdrücke erzeugen genau die regulären Sprachen.

Beweis in zwei Teilen:

1. Jede von einem regulären Ausdruck erzeugte Sprache ist regulär.
2. Für jede reguläre Sprache gibt es einen regulären Ausdruck, der sie erzeugt.

Beweis: Satz von Kleene (1)

1. Jede von einem regulären Ausdruck erzeugte Sprache ist regulär.
   Beweis:

Beweis: Satz von Kleene (1)

1. Jede von einem regulären Ausdruck erzeugte Sprache ist regulär.

Beweis:

- Wir konstruieren für regulären Ausdruck α einen NFA M_α mit ε-Übergängen und eindeutigen Start- und Endzuständen, sodass $L(M_\alpha) = L(\alpha)$.

Beweis: Satz von Kleene (1)

1. Jede von einem regulären Ausdruck erzeugte Sprache ist regulär.

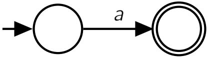
Beweis:

- Wir konstruieren für regulären Ausdruck α einen NFA M_α mit ε-Übergängen und eindeutigen Start- und Endzuständen, sodass $L(M_\alpha) = L(\alpha)$.
- Induktion über die Struktur von α.

Beweis: Satz von Kleene (1)

1. Jede von einem regulären Ausdruck erzeugte Sprache ist regulär.

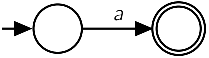
Beweis:

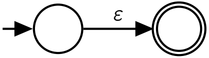
- Wir konstruieren für regulären Ausdruck α einen NFA M_α mit ε-Übergängen und eindeutigen Start- und Endzuständen, sodass $L(M_\alpha) = L(\alpha)$.
- Induktion über die Struktur von α.
- Basisfälle:
  - Für $\alpha = a \in \Sigma$ konstruiere 

Beweis: Satz von Kleene (1)

1. Jede von einem regulären Ausdruck erzeugte Sprache ist regulär.

Beweis:

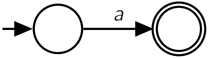
- Wir konstruieren für regulären Ausdruck α einen NFA M_α mit ε-Übergängen und eindeutigen Start- und Endzuständen, sodass $L(M_\alpha) = L(\alpha)$.
- Induktion über die Struktur von α.
- Basisfälle:
  - Für $\alpha = a \in \Sigma$ konstruiere 

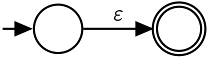
  - Für $\alpha = \varepsilon$ konstruiere 

Beweis: Satz von Kleene (1)

1. Jede von einem regulären Ausdruck erzeugte Sprache ist regulär.

Beweis:

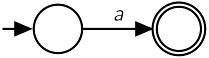
- Wir konstruieren für regulären Ausdruck α einen NFA M_α mit ε-Übergängen und eindeutigen Start- und Endzuständen, sodass $L(M_\alpha) = L(\alpha)$.
- Induktion über die Struktur von α.
- Basisfälle:
  - Für $\alpha = a \in \Sigma$ konstruiere 

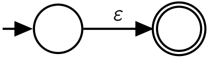
  - Für $\alpha = \varepsilon$ konstruiere 

  - Für $\alpha = \emptyset$ konstruiere 

Beweis: Satz von Kleene (1)

1. Jede von einem regulären Ausdruck erzeugte Sprache ist regulär.
Beweis:

- Wir konstruieren für regulären Ausdruck α einen NFA M_α mit ε-Übergängen und eindeutigen Start- und Endzuständen, sodass $L(M_\alpha) = L(\alpha)$.
- Induktion über die Struktur von α.
- Basisfälle:
  - Für $\alpha = a \in \Sigma$ konstruiere 

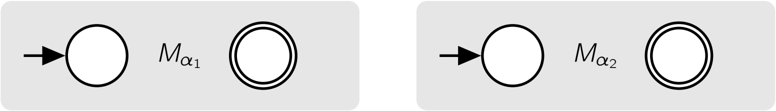
  - Für $\alpha = \varepsilon$ konstruiere 

  - Für $\alpha = \emptyset$ konstruiere 

In allen Fällen ist $L(M_\alpha) = L(\alpha)$ offensichtlich.

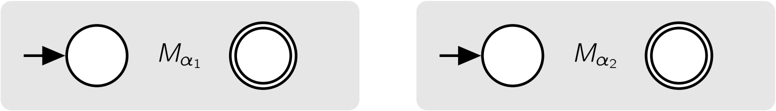
Beweis: Satz von Kleene (2)

- Induktionsschritt: Betrachte den Aufbau von α (3 Fälle)
  - Für $\alpha = \alpha_1\alpha_2$, liefert die I.H. $M_{\alpha_1}, M_{\alpha_2}$.

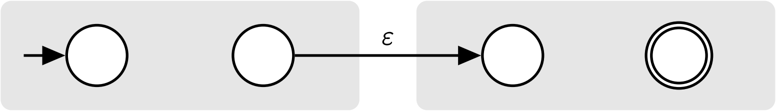


Beweis: Satz von Kleene (2)

- Induktionsschritt: Betrachte den Aufbau von α (3 Fälle)
  - Für $\alpha = \alpha_1\alpha_2$, liefert die I.H. $M_{\alpha_1}, M_{\alpha_2}$.

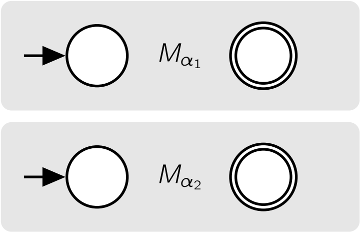


Konstruiere daraus M_α:



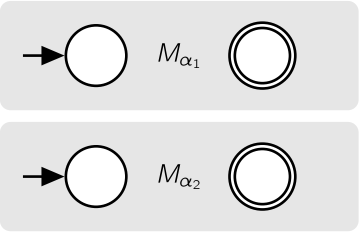
Beweis: Satz von Kleene (3)

- Für $\alpha = (\alpha_1|\alpha_2)$ liefert die I.H. $M_{\alpha_1}, M_{\alpha_2}$:

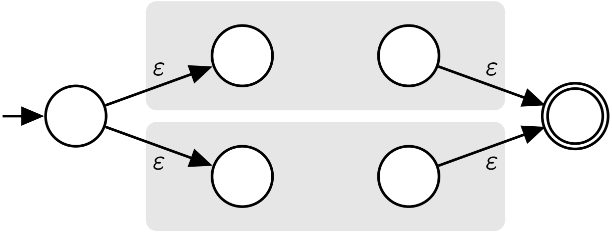


Beweis: Satz von Kleene (3)

- Für $\alpha = (\alpha_1|\alpha_2)$ liefert die I.H. $M_{\alpha_1}, M_{\alpha_2}$:

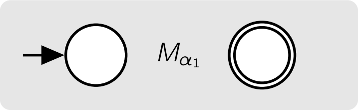


Konstruiere daraus M_α:



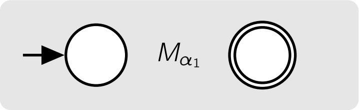
Beweis: Satz von Kleene (4)

- Für $\alpha = (\alpha_1)^*$ liefert die I.H. M_{α_1}

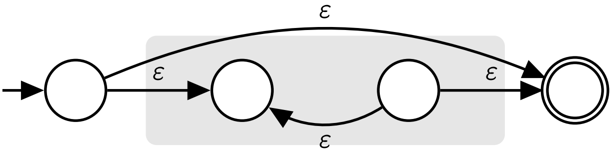


Beweis: Satz von Kleene (4)

- Für $\alpha = (\alpha_1)^*$ liefert die I.H. M_{α_1}



Konstruiere daraus M_α:



Einschub, Beispiel: Regulärer Ausdruck $\rightarrow$ NFA mit ε-Übergängen

NFA zum regulären Ausdruck

$$(\varepsilon|(a|b)^*b(a|b))$$

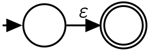
konstruieren:

Einschub, Beispiel: Regulärer Ausdruck $\to$ NFA mit ε-Übergängen

NFA zum regulären Ausdruck

$$(\varepsilon|(a|b)^* b(a|b))$$

konstruieren:

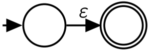


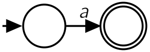
Einschub, Beispiel: Regulärer Ausdruck $\rightarrow$ NFA mit ε-Übergängen

NFA zum regulären Ausdruck

$$(\varepsilon|(a|b)^*b(a|b))$$

konstruieren:



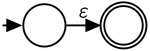


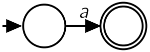
Einschub, Beispiel: Regulärer Ausdruck $\to$ NFA mit ε-Übergängen

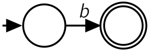
NFA zum regulären Ausdruck

$$(\varepsilon|(a|b)^*b(a|b))$$

konstruieren:





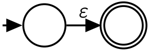


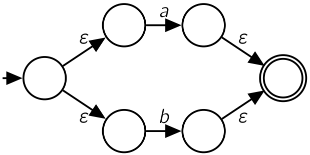
Einschub, Beispiel: Regulärer Ausdruck $\to$ NFA mit ε-Übergängen

NFA zum regulären Ausdruck

$$(\varepsilon|(a|b)^*b(a|b))$$

konstruieren:



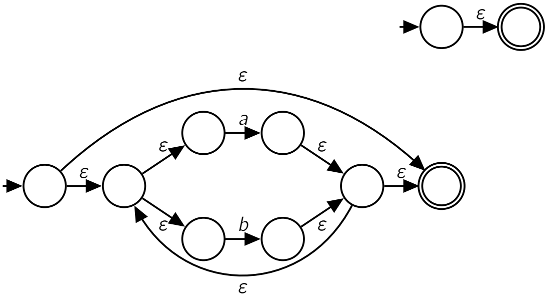


Einschub, Beispiel: Regulärer Ausdruck $\to$ NFA mit ε-Übergängen

NFA zum regulären Ausdruck

$$(\varepsilon|(a|b)^*b(a|b))$$

konstruieren:

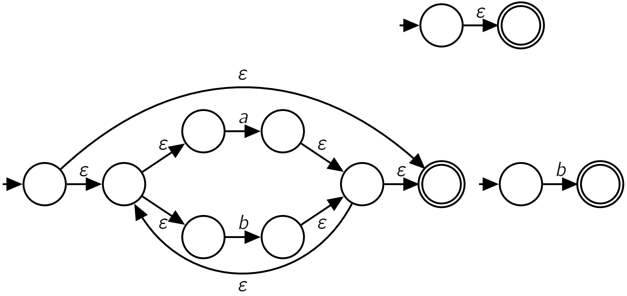


Einschub, Beispiel: Regulärer Ausdruck $\to$ NFA mit ε-Übergängen

NFA zum regulären Ausdruck

$$(\varepsilon|(a|b)^* \color{red}{b}\color{black}(a|b))$$

konstruieren:

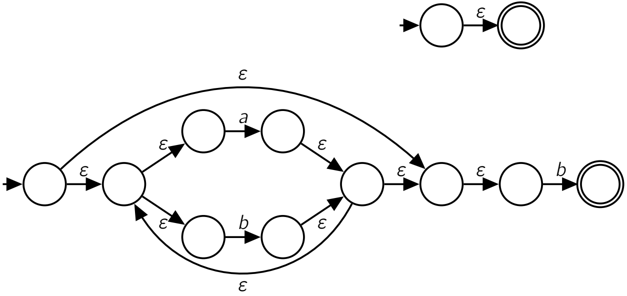


Einschub, Beispiel: Regulärer Ausdruck $\to$ NFA mit ε-Übergängen

NFA zum regulären Ausdruck

$$(\varepsilon|(a|b)^*b(a|b))$$

konstruieren:

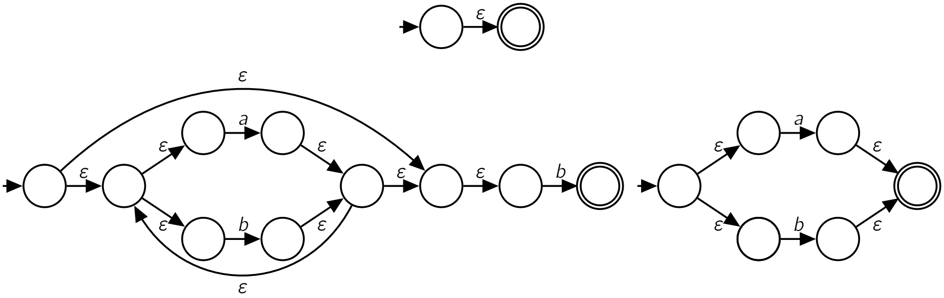


Einschub, Beispiel: Regulärer Ausdruck $\to$ NFA mit ε-Übergängen

NFA zum regulären Ausdruck

$$(\varepsilon|(a|b)^*b(a|b))$$

konstruieren:

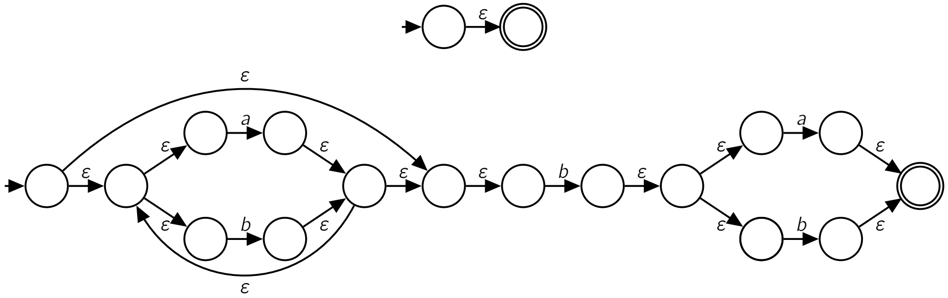


Einschub, Beispiel: Regulärer Ausdruck $\rightarrow$ NFA mit ε-Übergängen

NFA zum regulären Ausdruck

$$(\varepsilon|(a|b)^*b(a|b))$$

konstruieren:

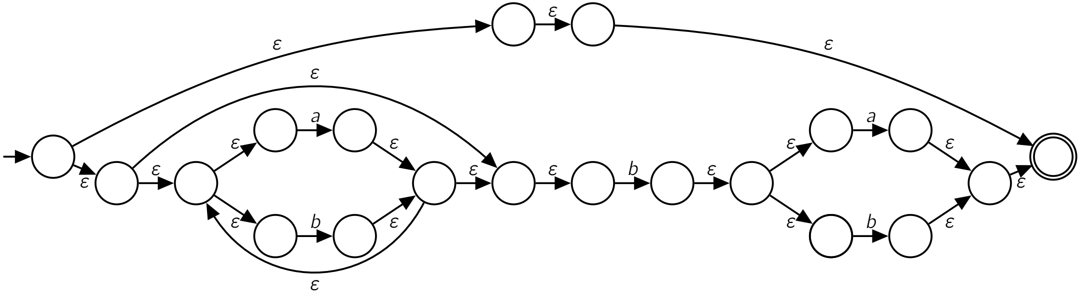


Einschub, Beispiel: Regulärer Ausdruck → NFA mit ε-Übergängen

NFA zum regulären Ausdruck

$$(\varepsilon|(a|b)^* b(a|b))$$

konstruieren:



Beweis: Satz von Kleene (5)

2. Für jede reguläre Sprache L gibt es einen regulären Ausdruck α mit $L(\alpha) = L$

Beweis:

Beweis: Satz von Kleene (5)

2. Für jede reguläre Sprache L gibt es einen regulären Ausdruck α mit $L(\alpha) = L$

Beweis:

- Sei DFA $M = (\{z_1, \ldots, z_n\}, \Sigma, \delta, z_1, E)$ mit $L(M) = L$ gegeben.

Beweis: Satz von Kleene (5)

2. Für jede reguläre Sprache L gibt es einen regulären Ausdruck α mit $L(\alpha) = L$

Beweis:

- Sei DFA $M = (\{z_1, \ldots, z_n\}, \Sigma, \delta, z_1, E)$ mit $L(M) = L$ gegeben.
- Für $w \in \Sigma^*$ und z_i, z_j mit $\widehat{\delta}(z_i, w) = z_j$ sei $visit_i(w) = q_1, \ldots, q_m$ die Folge der besuchten Zustände (wobei $q_1 = z_i$ und $q_m = z_j$).

Beweis: Satz von Kleene (5)

2. Für jede reguläre Sprache L gibt es einen regulären Ausdruck α mit $L(\alpha) = L$

Beweis:

- Sei DFA $M = (\{z_1, \ldots, z_n\}, \Sigma, \delta, z_1, E)$ mit $L(M) = L$ gegeben.
- Für $w \in \Sigma^*$ und z_i, z_j mit $\widehat{\delta}(z_i, w) = z_j$ sei $visit_i(w) = q_1, \ldots, q_m$ die Folge der besuchten Zustände (wobei $q_1 = z_i$ und $q_m = z_j$).
- Wir definieren:

$$L_{i,j}^k = \left\{ w \in \Sigma^* \;\middle|\; \begin{array}{l} \widehat{\delta}(z_i, w) = z_j \text{ und } visit_i(w) = q_1, \ldots, q_m, \\ \text{sodass für } 1 < l < m\text{: wenn } q_l = z_p \text{ dann } p \leq k \end{array} \right\}$$

$L_{i,j}^k$ enthält die Wörter, die von Zustand z_i zu Zustand z_j führen ohne dabei Zwischenzustände mit Index größer als k zu benutzen.

Beweis: Satz von Kleene (5)

2. Für jede reguläre Sprache L gibt es einen regulären Ausdruck α mit $L(\alpha) = L$

Beweis:

- Sei DFA $M = (\{z_1, \ldots, z_n\}, \Sigma, \delta, z_1, E)$ mit $L(M) = L$ gegeben.
- Für $w \in \Sigma^*$ und z_i, z_j mit $\widehat{\delta}(z_i, w) = z_j$ sei $\mathit{visit}_i(w) = q_1, \ldots, q_m$ die Folge der besuchten Zustände (wobei $q_1 = z_i$ und $q_m = z_j$).
- Wir definieren:

$$L^k_{i,j} = \left\{ w \in \Sigma^* \;\middle|\; \begin{array}{l} \widehat{\delta}(z_i, w) = z_j \text{ und } \mathit{visit}_i(w) = q_1, \ldots, q_m, \\ \text{sodass für } 1 < l < m\text{: wenn } q_l = z_p \text{ dann } p \leq k \end{array} \right\}$$

$L^k_{i,j}$ enthält die Wörter, die von Zustand z_i zu Zustand z_j führen ohne dabei Zwischenzustände mit Index größer als k zu benutzen.

- Mit Induktion über k zeigen wir:

$$\text{Es gibt reguläre Ausdrücke } \alpha^k_{i,j} \text{ mit } L(\alpha^k_{i,j}) = L^k_{i,j}.$$

Beweis: Satz von Kleene (6)

Zur Erinnerung: $L_{i,j}^k$ enthält die Wörter, die von Zustand z_i zu Zustand z_j führen ohne dabei Zwischenzustände mit Index größer als k zu benutzen.

Basis: $k = 0$.

- Wenn $i \neq j$, dann ist $L_{i,j}^0 = \{a \in \Sigma \mid \delta(z_i, a) = z_j\}$.

  Falls $L_{i,j}^0 = \{a_1, \ldots, a_q\}$ mit $q > 0$, dann gilt $L(\alpha_{i,j}^0) = L_{i,j}^0$ für $\alpha_{i,j}^0 = (a_1 | \ldots | a_q)$.

  Falls $L_{i,j}^0 = \emptyset$, dann gilt $L(\alpha_{i,j}^0) = L_{i,j}^0$ mit $\alpha_{i,j}^0 = \emptyset$.

Beweis: Satz von Kleene (6)

Zur Erinnerung: $L^k_{i,j}$ enthält die Wörter, die von Zustand z_i zu Zustand z_j führen ohne dabei Zwischenzustände mit Index größer als k zu benutzen.

Basis: $k = 0$.

- Wenn $i \neq j$, dann ist $L^0_{i,j} = \{a \in \Sigma \mid \delta(z_i, a) = z_j\}$.

  Falls $L^0_{i,j} = \{a_1, \ldots, a_q\}$ mit $q > 0$, dann gilt $L(\alpha^0_{i,j}) = L^0_{i,j}$ für $\alpha^0_{i,j} = (a_1| \ldots |a_q)$.

  Falls $L^0_{i,j} = \emptyset$, dann gilt $L(\alpha^0_{i,j}) = L^0_{i,j}$ mit $\alpha^0_{i,j} = \emptyset$.

- Wenn $i = j$, dann ist $L^0_{i,i} = \{\varepsilon\} \cup \{a \in \Sigma \mid \delta(z_i, a) = z_i\}$.

  Sei $L^0_{i,i} = \{\varepsilon, a_1, \ldots, a_q\}$.

  Dann gilt $L(\alpha^0_{i,i}) = L^0_{i,i}$ für $\alpha^0_{i,i} = (\varepsilon|a_1| \ldots |a_q)$.

Beweis: Satz von Kleene (7)

Zur Erinnerung: $L^k_{i,j}$ enthält die Wörter, die von Zustand z_i zu Zustand z_j führen ohne dabei Zwischenzustände mit Index größer als k zu benutzen.

Induktionsschritt: $k \to k+1$.

$$L^{k+1}_{i,j} = L^k_{i,j} \cup L^k_{i,k+1}(L^k_{k+1,k+1})^* L^k_{k+1,j}$$

denn entweder läuft M ohne Zustand z_{k+1} zu besuchen, oder der „Lauf“ kann in 3 Teile gespalten werden:

1. „Lauf“ von z_i bis zum ersten Besuch des Zustands z_{k+1} (abgedeckt durch $L^k_{i,k+1}$)
2. Mehrmaliges, zyklisches Besuchen von $k+1$ (beliebig oft) (abgedeckt durch $L^k_{k+1,k+1}$)
3. Letztmaliges Verlassen von z_{k+1} und „Lauf“ bis zu z_j (abgedeckt durch $L^k_{k+1,j}$)

Daher gilt $\alpha^{k+1}_{i,j} = (\alpha^k_{i,j} | \alpha^k_{i,k+1}(\alpha^k_{k+1,k+1})^* \alpha^k_{k+1,j})$ und $L(\alpha^{k+1}_{i,j}) = L^{k+1}_{i,j}$.

Beweis: Satz von Kleene (8)

Zur Erinnerung: $L_{i,j}^k$ enthält die Wörter, die von Zustand z_i zu Zustand z_j führen ohne dabei Zwischenzustände mit Index größer als k zu benutzen.

Damit haben wir bewiesen:

Es gibt reguläre Ausdrücke $\alpha_{i,j}^k$ mit $L(\alpha_{i,j}^k) = L_{i,j}^k$.

Jetzt müssen wir noch zeigen, dass es einen regulären Ausdruck gibt, der $L(M)$ erzeugt.

Beweis: Satz von Kleene (8)

Zur Erinnerung: $L_{i,j}^k$ enthält die Wörter, die von Zustand z_i zu Zustand z_j führen ohne dabei Zwischenzustände mit Index größer als k zu benutzen.

Damit haben wir bewiesen:

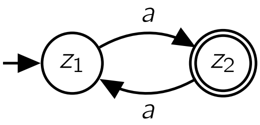
Es gibt reguläre Ausdrücke $\alpha_{i,j}^k$ mit $L(\alpha_{i,j}^k) = L_{i,j}^k$.

Jetzt müssen wir noch zeigen, dass es einen regulären Ausdruck gibt, der $L(M)$ erzeugt.

Sei die Menge der Endzustände $E = \{z_{i_1}, \ldots, z_{i_r}\}$.

Dann gilt $L(M) = \bigcup_{z_i \in E} L_{1,i}^n = L_{1,i_1}^n \cup \cdots \cup L_{1,i_r}^n = L(\alpha_{1,i_1}^n | \alpha_{1,i_2}^n | \ldots | \alpha_{1,i_r}^n)$. $\square$

Beispiel: DFA $\rightarrow$ regulärer Ausdruck



Regulärer Ausdruck dazu:

$$\begin{aligned}\alpha^2_{1,2} &= (\alpha^1_{1,2}|\alpha^1_{1,2}(\alpha^1_{2,2})^*\alpha^1_{2,2})\\ &= \big((a|\varepsilon(\varepsilon)^*a)\,|\,(a|\varepsilon(\varepsilon)^*a)\,(\varepsilon|a(\varepsilon)^*a)^*\,(\varepsilon|a(\varepsilon)^*a)\big)\end{aligned}$$

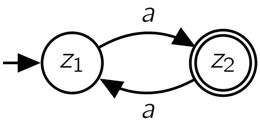
denn

$$\alpha^0_{1,1} = \varepsilon \qquad \alpha^0_{2,2} = \varepsilon \qquad \alpha^0_{1,2} = a \qquad \alpha^0_{2,1} = a$$

$$\alpha^1_{1,2} = (\alpha^0_{1,2}|\alpha^0_{1,1}(\alpha^0_{1,1})^*\alpha^0_{1,2}) = (a|\varepsilon(\varepsilon)^*a)$$

$$\alpha^1_{2,2} = (\alpha^0_{2,2}|\alpha^0_{2,1}(\alpha^0_{1,1})^*\alpha^0_{1,2}) = (\varepsilon|a(\varepsilon)^*a)$$

Beispiel: DFA $\to$ regulärer Ausdruck



Regulärer Ausdruck dazu:

$$\begin{aligned}\alpha^2_{1,2} &= (\alpha^1_{1,2}|\alpha^1_{1,2}(\alpha^1_{2,2})^*\alpha^1_{2,2}) \\ &= \big((a|\varepsilon(\varepsilon)^*a)\,|\,(a|\varepsilon(\varepsilon)^*a)\,(\varepsilon|a(\varepsilon)^*a)^*\,(\varepsilon|a(\varepsilon)^*a)\big)\end{aligned}$$

äquivalent zu $a(aa)^*$ (durch Vereinfachung)

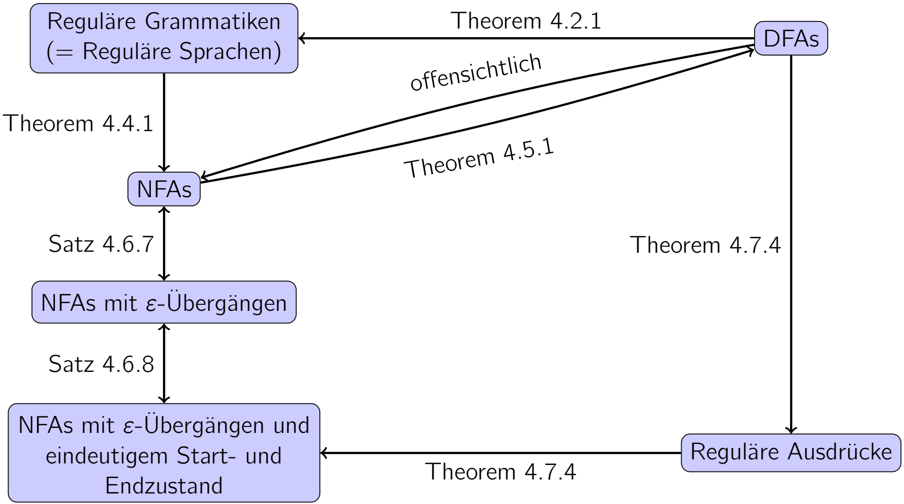
denn

$$\alpha^0_{1,1} = \varepsilon \qquad \alpha^0_{2,2} = \varepsilon \qquad \alpha^0_{1,2} = a \qquad \alpha^0_{2,1} = a$$

$$\alpha^1_{1,2} = (\alpha^0_{1,2}|\alpha^0_{1,1}(\alpha^0_{1,1})^*\alpha^0_{1,2}) = (a|\varepsilon(\varepsilon)^*a)$$

$$\alpha^1_{2,2} = (\alpha^0_{2,2}|\alpha^0_{2,1}(\alpha^0_{1,1})^*\alpha^0_{1,2}) = (\varepsilon|a(\varepsilon)^*a)$$

Zusammenfassung: Formalismen für reguläre Sprachen



Kanten ($\rightarrow$) zeigen Kodierungen