Theoretische Informatik 1

WiSe 23/24

Prof. Dr. Sebastian Siebertz
AG Theoretische Informatik
MZH, Raum 3160
siebertz@uni-bremen.de

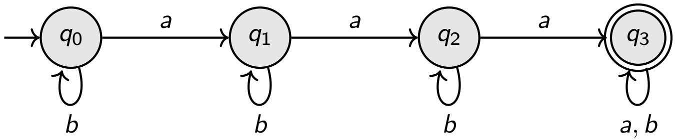
Universität Bremen

Wiederholung

- Alle (sinnvollen) Ein- und Ausgaben können als Wörter über Alphabeten kodiert werden.
- Ein Alphabet ist eine endliche, nichtleere Menge von Symbolen.
- Ein Wort über einem Alphabet Σ ist eine endliche Folge $w = w_1 \cdots w_n$ von Symbolen aus Σ.
- Σ^* ist die Menge aller Wörter über Σ.
- $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$.
- Eine Sprache über Σ ist eine Menge $L \subseteq \Sigma^*$.

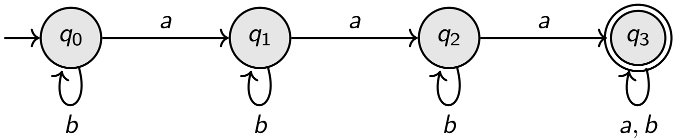
Wiederholung

- Automaten beschreiben formale Sprachen:
  - erhalten Wörter als Eingabe, lesen diese von links nach rechts,
  - befinden sich nach jedem Leseschritt in genau einem von endlich vielen möglichen Zuständen,
  - arbeiten ein „festes Programm“ ab,
  - verwerfen oder akzeptieren Eingabe über den am Ende erreichten Zustand.



Wiederholung

- Ein deterministischer endlicher Automat (DEA) ist ein Tupel $\mathcal{A} = (Q, \Sigma, q_s, \delta, F)$, wobei
  - Q eine endliche Menge von Zuständen ist,
  - Σ ein Eingabealphabet ist,
  - $q_s \in Q$ der Anfangszustand ist,
  - $\delta : Q \times \Sigma \to Q$ die Übergangsfunktion ist,
  - $F \subseteq Q$ eine Menge von akzeptierenden Zuständen ist.



Wiederholung

- Die kanonische Fortsetzung von $\delta\colon Q \times \Sigma \to Q$ von einzelnen Symbolen auf beliebige Wörter ist die Funktion $\hat{\delta} : Q \times \Sigma^* \to Q$, die induktiv (über die Wortlänge) definiert ist als:
  - $\hat{\delta}(q, \varepsilon) := q$
  - $\hat{\delta}(q, wa) := \delta(\hat{\delta}(q, w), a)$.
- Ein DEA $\mathcal{A} = (Q, \Sigma, q_s, \delta, F)$ akzeptiert ein Wort $w \in \Sigma^*$, wenn

$$\hat{\delta}(q_s, w) \in F.$$

  Die von $\mathcal{A}$ akzeptierte Sprache ist

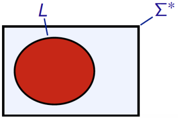
$$L(\mathcal{A}) = \{w \in \Sigma^* \mid \mathcal{A} \text{ akzeptiert } w\}.$$

- Eine Sprache $L \subseteq \Sigma^*$ heißt regulär, wenn es einen DEA $\mathcal{A}$ gibt mit $L = L(\mathcal{A})$.

Fragestellung

Welche Eigenschaften haben reguläre Sprachen?

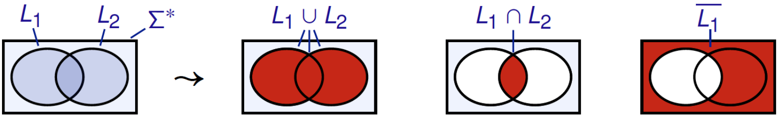
Operationen auf Wörtern und Sprachen



- Boolesche Operationen:

  … die üblichen Booleschen Mengenoperationen, angewendet auf Sprachen:

  - Vereinigung: $L_1 \cup L_2 = \{w : w \in L_1$ oder $w \in L_2\}$.
  - Schnitt: $L_1 \cap L_2 = \{w : w \in L_1$ und $w \in L_2\}$.
  - Komplement: $\bar{L}_1 = \{w : w \in \Sigma^*$ und $w \notin L_1\}$.



Operationen auf Wörtern und Sprachen

- Beispiele:
  - Wenn $L_1 = \{ab, ac\}$ und $L_2 = \{ba, ca\}$,

    dann $L_1 \cup L_2 = \{ab, ac, ba, ca\}$,

    und $L_1 \cap L_2 = \varnothing$.
  - Wenn $L_1 = \{w \in \Sigma^* : w \text{ hat gerade Länge}\}$,

    dann $\bar{L}_1 = \{w \in \Sigma^* : w \text{ hat ungerade Länge}\}$.

Operationen auf Wörtern und Sprachen

- Kleene-Stern:

  ... beliebig (aber nur endlich) oft iterierte Konkatenation.

  - Wir definieren zunächst die unendliche Sprachfolge $L^0, L^1, L^2, \ldots$
    - $L^0 := \{\varepsilon\}$
    - $L^{n+1} := L^n \cdot L$ für alle $n \geq 0$.
  - Kleene-Stern: $L^* := \bigcup_{n \geq 0} L^n$.
- Beispiel:
  - Wenn $L = \{ab, ba\}$, dann
    - $L^0 = \{\varepsilon\}$
    - $L^1 = \{\varepsilon\} \cdot L = \{ab, ba\}$
    - $L^2 = \{ab, ba\} \cdot L = \{abab, abba, baab, baba\}$
    - $\ldots$
    - $L^* = \{\varepsilon, ab, ba, abab, abba, \ldots\}$.

Operationen auf Wörtern und Sprachen

- Kleene-Stern, anders formuliert:
  - $L^* = \{\varepsilon\} \cup \{w : \text{es ex. } u_1, u_2, \ldots, u_n \in L, \text{ so dass } w = u_1 \cdot u_2 \cdot \ldots \cdot u_n\}$.
- Rechengesetze
  - $(L^*)^* = L^*$.
  - $L^* \cdot L^* = L^*$.
- Variante ohne das leere Wort
  - $L^+ := L^* \setminus \{\varepsilon\}$.

Operationen auf Wörtern und Sprachen

- Zusätzliche Notation:
  - $|w|$ bezeichnet die Länge des Wortes w (= Anzahl Symbole)

    z.B. $|aabba| = 5$, $|\varepsilon| = 0$.
  - $|w|_a$ für $a \in \Sigma$ bezeichnet die Anzahl Vorkommen des Symbols a im Wort w,

    z.B. $|abbaa|_a = 3$, $|abbaa|_b = 2$.
  - a^n bezeichnet das Wort, das aus n-mal dem Symbol a besteht,

    z.B. $a^3 = aaa$
  - w^n bezeichnet das Wort, das aus n-mal dem Wort w besteht,

    z.B. $(abc)^3 = abcabcabc$, aber $abc^3 = abccc$.

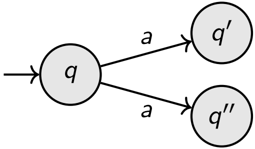
Operationen auf Wörtern und Sprachen

- Präfix, Infix, Suffix:
  - u ist Präfix von v wenn $v = uw$ für ein $w \in \Sigma^*$.
  - u ist Suffix von v wenn $v = wu$ für ein $w \in \Sigma^*$.
  - u ist Infix von v wenn $v = w_1 u w_2$ für ein $w_1, w_2 \in \Sigma^*$.

    Statt Infix sagen wir auch Teilwort.
- Beispiele:
  - ab is Präfix von $abba$.
  - a is Suffix von $abba$.
  - b ist Infix von $abba$.

Nichtdeterministische endliche Automaten

- Wir generalisieren nun unser Automatenmodell, indem wir auch Nichtdeterminismus erlauben:



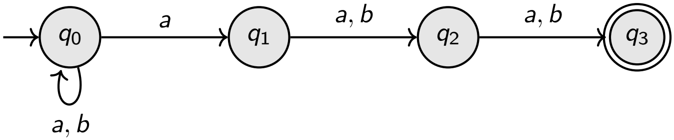
- Ein Automat hat nun mehrere Möglichkeiten, ein Wort zu verarbeiten.
- Er akzeptiert eine Eingabe, wenn eine Möglichkeit existiert, einen Endzustand zu erreichen.
- Nichtdeterminismus ist ein fundamentales Konzept der Informatik; wir werden es noch häufig verwenden!

Nichtdeterministische endliche Automaten

Michael O. Rabin, Dana Scott

- Ein nichtdeterministischer endlicher Automat (NEA) ist ein Tupel $\mathcal{A} = (Q, \Sigma, q_s, \Delta, F)$, wobei
  - Q eine endliche Menge von Zuständen ist,
  - Σ ein Eingabealphabet ist,
  - $q_s \in Q$ der Anfangszustand ist,
  - $\Delta \subseteq Q \times \Sigma \times Q$ die Übergangsrelation ist,
  - $F \subseteq Q$ eine Menge von akzeptierenden Zuständen ist.

Beispiel



Dieser Automat ist kein DEA, da es q_0 für a zwei mögliche Übergänge gibt und an Stelle q_3 keine möglichen Übergänge.

Akzeptanz von NEAs

- Ein Lauf in einem NEA $\mathcal{A} = (Q, \Sigma, q_s, \Delta, F)$ von einem Zustand $p_0 \in Q$ zu einem Zustand $p_n \in Q$ ist eine Folge

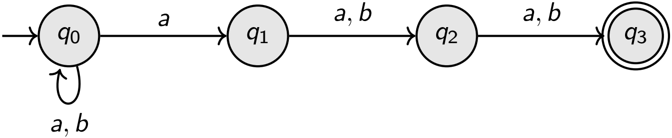
$$\pi \;=\; p_0 \xrightarrow{a_1}_{\mathcal{A}} p_1 \xrightarrow{a_2}_{\mathcal{A}} \cdots \xrightarrow{a_n}_{\mathcal{A}} p_n$$

  mit $(p_i, a_{i+1}, p_{i+1}) \in \Delta$ für alle $0 \le i < n$.
- Der Lauf π hat die Beschriftung $w := a_1 \cdots a_n$.
- Wenn es in $\mathcal{A}$ einen Lauf von $p \in Q$ nach $q \in Q$ mit der Beschriftung w gibt, so schreiben wir:

$$p \overset{w}{\Longrightarrow}_{\mathcal{A}} q$$

- Für $n = 0$ sprechen wir vom leeren Lauf, der die Beschriftung ε hat.

Akzeptanz von NEAs: Beispiel

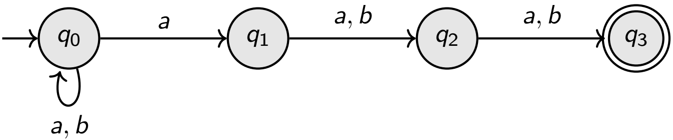


In diesem NEA gibt es unter anderem folgende Läufe für die Eingabe *aba*:

$$
\begin{aligned}
\pi_1 &= q_0 \xrightarrow{a}_{\mathcal{A}} q_1 \xrightarrow{b}_{\mathcal{A}} q_2 \xrightarrow{a}_{\mathcal{A}} q_3 \\
\pi_2 &= q_0 \xrightarrow{a}_{\mathcal{A}} q_0 \xrightarrow{b}_{\mathcal{A}} q_0 \xrightarrow{a}_{\mathcal{A}} q_1
\end{aligned}
$$

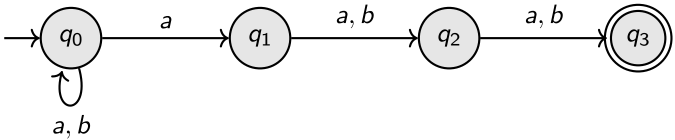
Akzeptanz von NEAs

- Der NEA $\mathcal{A} = (Q, \Sigma, q_s, \Delta, F)$ akzeptiert das Wort $w \in \Sigma^*$, wenn $q_s \xRightarrow{w}_{\mathcal{A}} q_f$ für ein $q_f \in F$.
- Die von $\mathcal{A}$ akzeptierte Sprache ist $L(\mathcal{A}) = \{w \in \Sigma^* \mid \mathcal{A} \text{ akzeptiert } w\}$.
- Beispiel:



  - Das Wort *aba* wird akzeptiert, weil ein akzeptierender Lauf existiert.
  - $L(\mathcal{A}) = \{w \in \{a, b\}^* \mid \text{das drittletzte Symbol in } w \text{ ist } a\}$.

Akzeptanz von NEAs



- $L(\mathcal{A}) = \{w \in \{a,b\}^* \mid \text{das drittletzte Symbol in } w \text{ ist } a\}$.
- Intuitionen:
  - Wenn $\mathcal{A}$ in q_0 ein a liest, so „rät“ er, ob er an der drittletzten Stelle ist.
  - Hat er „ja“ geraten, so verifiziert er mit der $q_1 - q_3$-Kette diese Wahl.
- Wichtig:
  - für Wörter $w \in L$ gibt es die Möglichkeit, richtig zu raten und
  - für Wörter $w \notin L$ führt falsches Raten niemals zur Akzeptanz.
- Es ist gar nicht so einfach, einen DEA für diese Sprache zu finden!

DEA vs NEA

- Offensichtlich ist jeder DEA auch ein NEA:
  - DEA-Übergang $\delta(q, a) = q'$ kann durch NEA-Übergang $(q, a, q') \in \Delta$ dargestellt werden.
- Damit ist jede durch einen DEA erkennbare Sprache auch durch einen NEA erkennbar.
- Gilt die Umkehrung auch?
  - Ja! DEAs und NEAs akzeptieren dieselben Sprachen.
  - Beweis mittels Potenzmengenkonstruktion.
  - Zwei NEAs heißen äquivalent, wenn sie dieselbe Sprache erkennen.

Der Satz von Rabin und Scott

Satz von Rabin und Scott

Zu jedem NEA kann ein äquivalenter DEA konstruiert werden.

Beweis:

- Sei $\mathcal{A}(Q, \Sigma, q_s, \Delta, F)$ ein NEA.
- Definiere $\mathcal{A}' = (2^Q, \Sigma, \{q_s\}, \delta, F')$ mit:
  - $\delta(P, a) = \bigcup_{p \in P} \{q \mid (p, a, q) \in \Delta\}$ für alle $P \in 2^Q$ und $a \in \Sigma$
  - $F' = \{P \in 2^Q \mid P \cap F \neq \varnothing\}$

Beispiel: Tafel

Der Satz von Rabin und Scott

- Hilfsaussage: $q \in \hat{\delta}(\{q_s\}, w) \quad \Leftrightarrow \quad q_s \xRightarrow{w}_{\mathcal{A}} q$.
- Beweis per Induktion über $|w|$:
  - Induktionsanfang: $|w| = 0$

    $$q \in \hat{\delta}(\{q_s\}, \varepsilon) \quad \Leftrightarrow \quad q \in \{q_s\} \quad \Leftrightarrow \quad q_s = q \quad \Leftrightarrow \quad q_s \xRightarrow{\varepsilon}_{\mathcal{A}} q \ \checkmark$$

  - Induktionsannahme: Die Hilfsaussage gilt für alle $w \in \Sigma^*$ mit $|w| \leq n$.

Der Satz von Rabin und Scott

- Hilfsaussage: $q \in \hat{\delta}(\{q_s\}, w) \quad \Leftrightarrow \quad q_s \stackrel{w}{\Longrightarrow}_{\mathcal{A}} q$.
  - Induktionsschritt: $|w| = n + 1$

    Sei $w = ua$ mit $u \in \Sigma^*$, $|u| = n$ und $a \in \Sigma$. Es gilt:

$$
\begin{aligned}
\hat{\delta}(\{q_s\}, ua) &= \delta(\hat{\delta}(\{q_s\}, u), a) && (\text{Def. } \hat{\delta}) \\
&= \bigcup_{p \in \hat{\delta}(\{q_s\}, u)} \{q \mid (p, a, q) \in \Delta\} && (\text{Def. } \delta) \\
&= \bigcup_{q_s \stackrel{u}{\Longrightarrow}_{\mathcal{A}} p} \{q \mid (p, a, q) \in \Delta\} && (\text{IV}) \\
&= \{q \mid q_s \stackrel{ua}{\Longrightarrow}_{\mathcal{A}} q\} && (\text{Def. Lauf})
\end{aligned}
$$

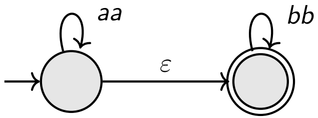
Beweis des Satzes von Rabin und Scott

Wir zeigen nun $L(\mathcal{A}) = L(\mathcal{A}')$:

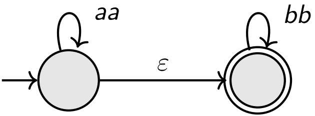
$$\begin{array}{rcll} w \in L(\mathcal{A}) & \Leftrightarrow & \exists q \in F : q_s \overset{w}{\Longrightarrow}_{\mathcal{A}} q & \text{(Def. } L(\mathcal{A})\text{)} \\ & \Leftrightarrow & \exists q \in F : q \in \hat{\delta}(\{q_s\}, w) & \text{(Hilfsaussage)} \\ & \Leftrightarrow & \hat{\delta}(\{q_s\}, w) \cap F \neq \varnothing & \\ & \Leftrightarrow & \hat{\delta}(\{q_s\}, w) \in F' & \text{(Def. } F'\text{)} \\ & \Leftrightarrow & w \in L(\mathcal{A}'). & \end{array}$$

Endliche Automaten mit Wortübergängen

- Ein NEA mit Wortübergängen hat die Form $\mathcal{A} = (Q, \Sigma, q_s, \Delta, F)$, wobei Q, Σ, q_s, F wie beim NEA definiert sind und $\Delta \subseteq Q \times \Sigma^* \times Q$ eine endliche Menge von Wortübergängen ist.
- Ein ε-NEA ist ein NEA mit Wortübergängen der Form (q, ε, q') und (q, a, q') mit $a \in \Sigma$.



Beispiel



- Läufe, Laufbeschriftungen und erkannte Sprache werden entsprechend wie für NEAs definiert.
- Zum Beispiel hat der Lauf

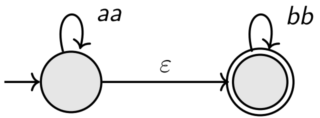
$$q_0 \xrightarrow{aa}_{\mathcal{A}} q_1 \xrightarrow{\varepsilon}_{\mathcal{A}} q_2 \xrightarrow{bb}_{\mathcal{A}} q_3$$

des Automaten die Beschriftung $aa \cdot \varepsilon \cdot bb = aabb$.

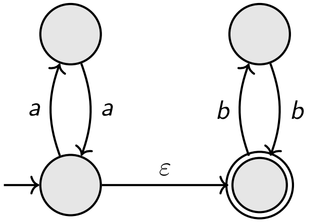
Äquivalenz der Modelle

Satz

Zu jedem NEA mit Wortübergängen kann ein äquivalenter ε-NEA konstruiert werden.



wird überführt in einen äquivalenten ε-NEA:



Äquivalenz der Modelle

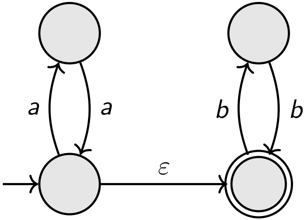
Satz

Zu jedem ε-NEA kann ein äquivalenter NEA konstruiert werden.

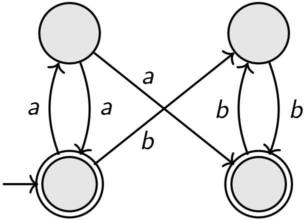
Beweis:

- Sei $\mathcal{A} = (Q, \Sigma, q_s, \Delta, F)$ ein ε-NEA.
- Definiere $\mathcal{A}' = (Q, \Sigma, q_s, \Delta', F')$, wobei
  - $\Delta' := \left\{ (p, a, q) \in Q \times \Sigma \times Q \mid p \overset{a}{\Longrightarrow}_{\mathcal{A}} q \right\}$
  - $F' := \begin{cases} F \cup \{q_s\} & \text{falls} \quad q_s \overset{\varepsilon}{\Longrightarrow}_{\mathcal{A}} q_f \text{ für ein } q_f \in F \\ F & \text{sonst} \end{cases}$
- $L(\mathcal{A}) = L(\mathcal{A}')$ siehe Skript.

Beispiel



wird in folgenden NEA überführt:



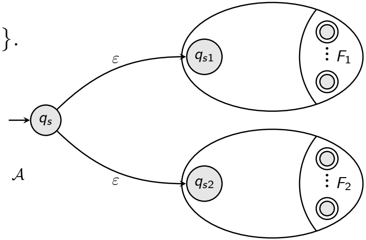
Abschluss unter Vereinigung (Ausblick)

Jetzt ist leicht zu zeigen: sind L_1 und L_2 regulär, so ist auch $L_1 \cup L_2$ regulär:

- Der folgende ε-NEA erkennt $L_1 \cup L_2$:

  $\mathcal{A} := (Q_1 \cup Q_2 \cup \{q_s\}, \Sigma, q_s, \Delta, F_1 \cup F_2)$, wobei

  - $q_s \notin Q_1 \cup Q_2$ und
  - $\Delta := \Delta_1 \cup \Delta_2 \cup \{(q_s, \varepsilon, q_{s1}), (q_s, \varepsilon, q_{s2})\}$.



- ε-NEA $\rightsquigarrow$ NEA ✓