Abschnitt 3

Sprachen

Unterabschnitt 1

Wörter

Wörter

Definition

- Ein Alphabet ist eine endliche Menge Σ von Zeichen.
- Ein Wort über Σ ist eine endliche Folge von Zeichen aus Σ.
- Σ^* ist die Menge der Wörter über Σ.

Beispiele

Sei $\Sigma = \{a, b, c\}$. Wörter über Σ:

- $ababbc$
- bac
- b
- das leere Wort, bezeichnet mit ε

Wörter

Frage

Warum verwendet man Wörter?

- Ausdrucksstärke
  Alles, was im Computer dargestellt werden kann, kann als Wort kodiert werden.
- mathematische Einfachheit

Operationen für Wörter

- Einbettung $\iota : \Sigma \to \Sigma^*$
  $\iota(a)$ auch kurz als a geschrieben
- Konkatenation $\cdot : \Sigma^{*2} \to \Sigma^*$
  $u \cdot v$ auch kurz als uv geschrieben
- Länge $|\cdot| : \Sigma^* \to \mathbb{N}$
  $|w|$
- n-tes Zeichen $\pi : D \to \Sigma$,
  wobei $D := \{(n, w) \mid n \in \mathbb{N} \setminus \{0\} \land w \in \Sigma^* \land n \leq |w|\}$
- n-te Potenz $\mathrm{pot} : \mathbb{N} \times \Sigma^* \to \Sigma^*$
  $\mathrm{pot}(n, w)$ kurz als w^n geschrieben
- Spiegelung $\cdot^R : \Sigma^* \to \Sigma^*$
  w^R

Unterabschnitt 2

Sprachen

Sprachen

Definition

Eine Sprache über einem Alphabet Σ ist eine Menge von Wörtern über Σ.

Beispiele

für Sprachen über dem Alphabet $\Sigma = \{a, b, c\}$:

- $\emptyset$
- Σ^*
- $\{a, abc, cbaa\}$
- $\{\varepsilon\}$
- $\{a\}$
- $\{a^n \mid n \in \mathbb{N}\} = \{\varepsilon, a, aa, aaa, \dots\}$

Sprachen

Beispiele

- $\Sigma := \{0, 1, +, *, (,)\}$.
  Die Sprache L sei die Menge der arithmetischen Ausdrücke über Σ.
- Σ sei die Menge der ASCII-Zeichen.
  Die Sprache L sei die Menge aller syntaktisch korrekten C-Programme.
- $\Sigma := \{1, x, y, z, +, *, =, (,)\}$.
  Die Sprache L sei die Menge der arithmetischen Gleichungen über Σ, die eine Lösung in $\mathbb{N}$ besitzen.
  - $(1 + 1) * (1 + 1 + 1) = 1 + 1 + 1 + 1 + 1 + 1 \ \in \ L$
  - $x + x = x * x \ \in \ L$
  - $(x + 1) * (x + 1) * (x + 1) + (y + 1) * (y + 1) * (y + 1) = z * z * z \ \notin \ L$

Verwendung von Sprachen

- als Mittel, um Informationen auszudrücken: Programmiersprachen, . . .
- Zur Formulierung von Entscheidungsproblemen: Gibt es einen Algorithmus, der entscheiden kann, ob ein gegebenes Wort in der Sprache liegt?

Operationen auf Sprachen

- Komplement $\overline{L} = \Sigma^* \setminus L$
- Vereinigung $L_1 \cup L_2$
- Durchschnitt $L_1 \cap L_2$
- Konkatenation $L_1 L_2 = \{uv \mid u \in L_1 \wedge v \in L_2\}$
- Kleene-Stern $L^* = \{w_1 \cdots w_n \mid n \in \mathbb{N} \wedge w_1, \ldots, w_n \in L\}$
- $L^+ = \{w_1 \cdots w_n \mid n \in \mathbb{N} \setminus \{0\} \wedge w_1, \ldots, w_n \in L\}$

Abschnitt 4

Reguläre Ausdrücke und reguläre Sprachen

Reguläre Ausdrücke

Definition

Sei Σ ein Alphabet.
Ein regulärer Ausdruck über Σ ist ein spezielles Wort über dem Alphabet $\Sigma \cup \{O, e, ^*, |, (,)\}$.
Induktive Definition:

- O ist ein regulärer Ausdruck.
- e ist ein regulärer Ausdruck.
- Für jedes Zeichen $a \in \Sigma$ ist das Wort a ein regulärer Ausdruck.
- Wenn A und B reguläre Ausdrücke sind, dann auch $(A|B)$.
- Wenn A und B reguläre Ausdrücke sind, dann auch (AB).
- Wenn A ein regulärer Ausdruck ist, dann auch A^*.

Reguläre Ausdrücke

Beispiele

- a
- (ab)
- $(a|b)$
- $(a^* b^*)$
- $(ab)^*$
- $(a^*|b^*)$
- $(a|b)^*$

Klammern werden oft weggelassen, wenn dies nicht zu Missverständnissen führt.

Die durch einen regulären Ausdruck bezeichnete Sprache

Definition

Die durch einen regulären Ausdruck A bezeichnete Sprache L_A ist folgendermaßen definiert.

- $L_0 = \emptyset$
- $L_e = \{\varepsilon\}$
- $L_a = \{a\}$ für $a \in \Sigma$
- $L_{(A|B)} = L_A \cup L_B$
- $L_{(AB)} = L_A L_B$
- $L_{A^*} = {L_A}^*$

Definition

Eine durch einen regulären Ausdruck bezeichnete Sprache heißt reguläre Sprache.

Die durch einen regulären Ausdruck bezeichnete Sprache

Beispiele

- $L_a = \{a\}$
- $L_{(ab)} = \{ab\}$
- $L_{(a|b)} = \{a, b\}$
- $L_{(a^* b^*)} = \{\varepsilon, b, bb, \ldots, a, ab, abb, \ldots\}$
- $L_{(ab)^*} = \{\varepsilon, ab, abab, \ldots\}$
- $L_{(a^*|b^*)} = \{\varepsilon, a, aa, \ldots, b, bb, \ldots\}$
- $L_{(a|b)^*} = \{\varepsilon, a, b, aa, ab, ba, bb, \ldots\}$

Reguläre Sprachen

Wenn L und L' reguläre Sprachen sind, so sind die folgenden Sprachen auch regulär.

- $L \cup L'$
- LL'
- L^*
- $\overline{L}$ (Beweis auf Umweg über endliche Automaten)
- $L \cap L'$ (Beweis auf Umweg über endliche Automaten)
- $L \setminus L'$ (folgt aus den vorigen beiden Aussagen)

Abschnitt 5

Endliche Automaten

Unterabschnitt 1

Deterministische endliche Automaten

Deterministische endliche Automaten

Definition

Ein deterministischer endlicher Automat ist ein Quintupel $(Q, \Sigma, \delta, q_0, F)$, das gegeben ist durch

- eine endliche Menge Q von Zuständen
- ein Alphabet Σ
- eine Übergangsfunktion $\delta\colon Q \times \Sigma \to Q$
- einen Anfangszustand (Startzustand) $q_0 \in Q$
- eine Menge $F \subseteq Q$ von Endzuständen

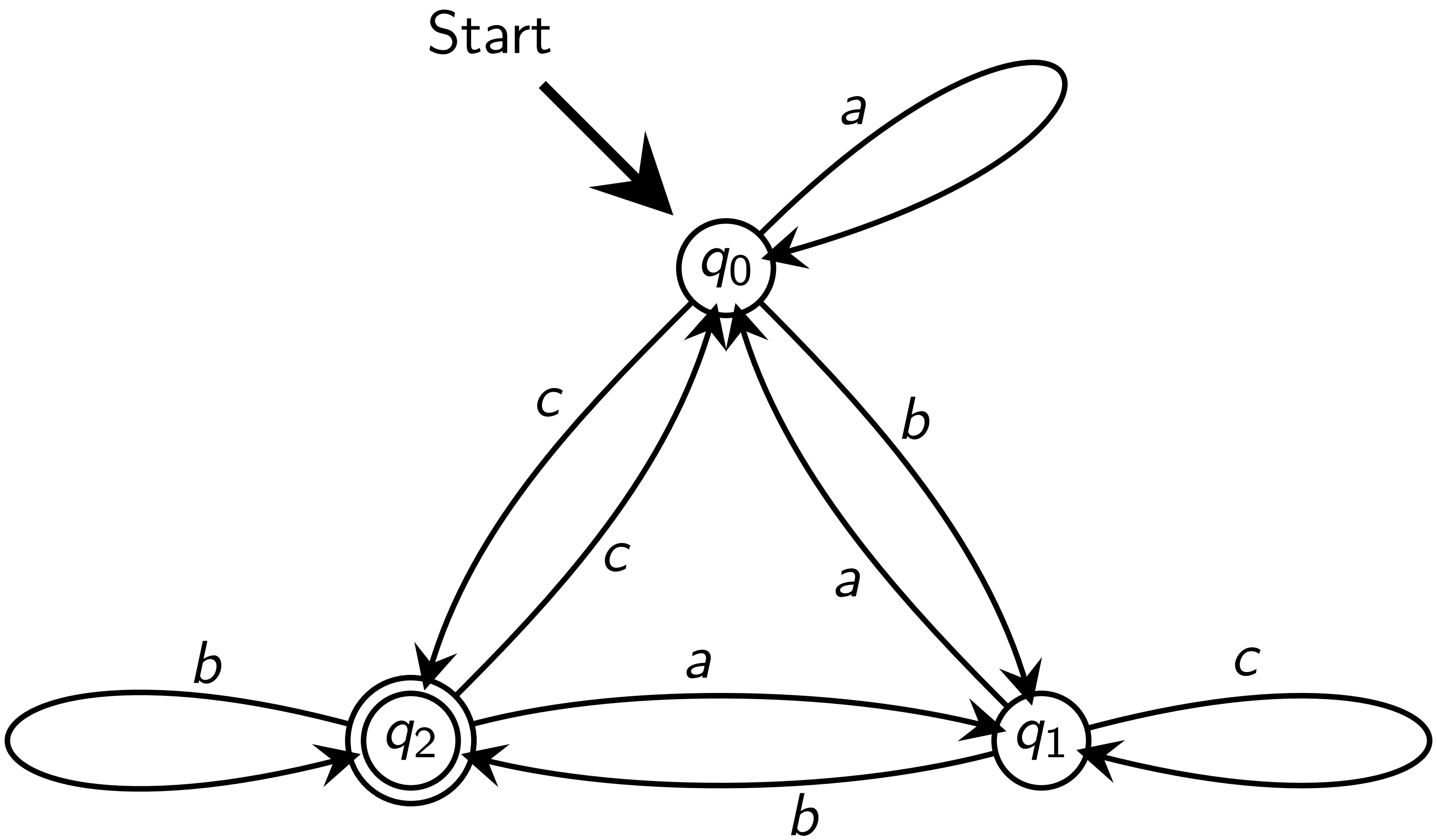
Deterministische endliche Automaten

Beispiel

Ein deterministischer endlicher Automat $(Q, \Sigma, \delta, q_0, F)$

- $Q = \{q_0, q_1, q_2\}$
- $\Sigma = \{a, b, c\}$
- $\delta(q_0, a) = q_0 \quad \delta(q_0, b) = q_1 \quad \delta(q_0, c) = q_2$
  $\delta(q_1, a) = q_0 \quad \delta(q_1, b) = q_2 \quad \delta(q_1, c) = q_1$
  $\delta(q_2, a) = q_1 \quad \delta(q_2, b) = q_2 \quad \delta(q_2, c) = q_0$
- $F = \{q_2\}$

Darstellung eines endlichen Automaten als Graph



Die von einem Automaten akzeptierte Sprache

Definition

Ein deterministischer endlicher Automat über einem Alphabet Σ akzeptiert ein Wort $w = a_1 \dots a_n$ aus Σ^*, wenn es eine endliche Folge $(q_0, \dots, q_n)$ von Zuständen gibt, sodass gilt

- q_0 ist der Anfangszustand des Automaten
- $q_i = \delta(q_{i-1}, a_i)$ für $i = 1, \dots, n$
- $q_n \in F$.

Definition

Die von einem deterministischen endlichen Automaten akzeptierte Sprache ist die Menge der von ihm akzeptierten Wörter.

Konfigurationen

Definition

Eine Konfiguration eines deterministischen endlichen Automaten $(Q, \Sigma, \delta, q_0, F)$ ist ein Paar (q, w), wobei $q \in Q$ und $w \in \Sigma^*$ ist.

Gibt den momentanen Zustand und den noch zu lesenden Teil des Eingabewortes an.

Berechnungen

Definition

Ein deterministischer endlicher Automat $(Q, \Sigma, \delta, q_0, F)$ führt eine Konfiguration (q, w) in eine Konfiguration (q', w') über, wenn es ein Zeichen $a \in \Sigma$ gibt mit

$$w = aw' \wedge q' = \delta(q, a).$$

Berechnungen

Definition

Eine Berechnung durch einen deterministischen endlichen Automaten $(Q, \Sigma, \delta, q_0, F)$ ist eine endliche Folge $(k_0, \ldots, k_n)$ von Konfigurationen derart, dass gilt:

- Der Automat führt die Konfiguration k_{i-1} in die Konfiguration k_i über für $i = 1, \ldots, n$.
- k_0 hat die Form (q_0, w).
- k_n hat die Form (q, ε).

Die Berechnung heißt akzeptierend, wenn $q \in F$ ist.

Von DEA akzeptierte Sprachen

Frage

Für welche Sprachen L gibt es einen deterministischen endlichen Automaten, der L akzeptiert?

Sei Σ ein Alphabet.

Satz

Die Menge der Sprachen L über Σ, zu denen es einen deterministischen endlichen Automaten gibt, der sie akzeptiert, ist abgeschlossen gegenüber Komplementbildung.

Definition

$L[u] := \{w \in \Sigma^* \mid uw \in L\}$ für $u \in \Sigma^*$

Von DEA akzeptierte Sprachen

Für einen deterministischen endlichen Automaten:

Definition

- q_u sei der Zustand, der nach Lesen des Wortes u erreicht ist.
- L_q sei die Menge aller Wörter, die den Automaten vom Zustand q aus in einen Endzustand überführen.

Lemma

Wenn ein deterministischer endlicher Automat die Sprache L akzeptiert, dann gilt für ihn $L[u] = L_{q_u}$ für alle $u \in \Sigma^*$.

Beweis.

$w \in L[u] \iff uw \in L \iff$ Der Automat akzeptiert $uw \iff w \in L_{q_u}$ $\square$

Von DEA akzeptierte Sprachen

$L[u] = L_{q_u}$.   Also $L[u] \in \{L_q \mid q \in Q\}$.

Satz

- *Es gibt genau dann einen deterministischen endlichen Automaten, der L akzeptiert, wenn es nur endlich viele verschiedene Sprachen $L[u]$ gibt.*
- *In diesem Fall lässt sich ein solcher deterministischer endlicher Automat mit einer minimalen Anzahl von Zuständen folgendermaßen konstruieren.*
  - *Zu jedem $L[u]$ gibt es einen Zustand q_u.*
  - *Ein Zeichen a führt vom Zustand q_u in den Zustand q_{ua}.*
  - *Anfangszustand ist q_ε.*
  - *Endzustände sind alle q_u mit $\varepsilon \in L[u]$.*
- *Er hat so viele Zustände, wie es Sprachen $L[u]$ gibt.*

Satz von Myhill-Nerode

Definition

Nerode-Relation $\sim_L$ einer Sprache $L \subseteq \Sigma^*$:

$$u \sim_L v :\iff \bigwedge_{w \in \Sigma^*} (uw \in L \iff vw \in L)$$

$$u \sim_L v \iff L[u] = L[v]$$

Satz (Myhill, Nerode)

- *Zu einer Sprache L existiert genau dann ein DEA, der sie akzeptiert, wenn der Index $|\Sigma^* / \sim_L|$ ihrer Nerode-Relation endlich ist.*
- *In diesem Fall ist die Anzahl der Zustände eines minimalen deterministischen endlichen Automaten, der L akzeptiert, gleich diesem Index.*

Unterabschnitt 2

Nichtdeterministische endliche Automaten

Nichtdeterministische endliche Automaten

Definition

Ein nichtdeterministischer endlicher Automat ist ein Quintupel $(Q, \Sigma, \Delta, q_0, F)$, das gegeben ist durch

- eine endliche Menge Q von Zuständen
- ein Alphabet Σ
- eine Übergangsrelation $\Delta \subseteq Q \times \Sigma \times Q$
- einen Anfangszustand (Startzustand) $q_0 \in Q$
- eine Menge $F \subseteq Q$ von Endzuständen

Die vom Automaten akzeptierte Sprache

Definition

Ein nichtdeterministischer endlicher Automat über einem Alphabet Σ akzeptiert ein Wort $w = a_1 \dots a_n$ aus Σ^*, wenn es eine endliche Folge $(q_0, \dots, q_n)$ von Zuständen gibt, sodass gilt

- q_0 ist der Anfangszustand des Automaten
- $(q_{i-1}, a_i, q_i) \in \Delta$ für $i = 1, \dots, n$
- $q_n \in F$.

Definition

Die von einem nichtdeterministischen endlichen Automaten akzeptierte Sprache ist die Menge der von ihm akzeptierten Wörter.

Konfigurationen

Definition

Eine Konfiguration eines nichtdeterministischen endlichen Automaten $(Q, \Sigma, \Delta, q_0, F)$ ist ein Paar (q, w), wobei $q \in Q$ und $w \in \Sigma^*$ ist.

Gibt den momentanen Zustand und den noch zu lesenden Teil des Eingabewortes an.

Berechnungen

Definition

Ein nichtdeterministischer endlicher Automat $(Q, \Sigma, \Delta, q_0, F)$ führt eine Konfiguration (q, w) in eine Konfiguration (q', w') über, wenn es ein Zeichen $a \in \Sigma$ gibt mit

$$w = aw' \wedge (q, a, q') \in \Delta.$$

Berechnungen

Definition

Eine Berechnung durch einen nichtdeterministischen endlichen Automaten $(Q, \Sigma, \Delta, q_0, F)$ ist eine endliche Folge $(k_0, \ldots, k_n)$ von Konfigurationen derart, dass gilt:

- Der Automat führt die Konfiguration k_{i-1} in die Konfiguration k_i über für $i = 1, \ldots, n$.
- k_0 hat die Form (q_0, w).
- k_n hat die Form (q, ε).

Die Berechnung heißt akzeptierend, wenn $q \in F$ ist.

Simulation eines nichtdeterministischen endlichen Automaten durch einen deterministischen

Gegeben ein nichtdeterministischer endlicher Automat $(Q, \Sigma, \Delta, q_0, F)$.

Konstruktion eines deterministischen endlichen Automaten, der die gleiche Sprache akzeptiert

Automat $(Q', \Sigma, \delta', q_0', F')$:

- $Q' = \mathcal{P}(Q)$
- $\delta'(p', a) = \{q \mid \bigvee_{p \in p'} (p, a, q) \in \Delta\}$
- $q_0' = \{q_0\}$
- $F' = \{R \subseteq Q \mid \bigvee_{r \in R} r \in F\}$

Unterabschnitt 3

Automaten mit Wort-Übergängen

Nichtdeterministische endliche Automaten mit Wortübergängen

Definition

Ein nichtdeterministischer endlicher Automat mit Wortübergängen ist ein Quintupel $(Q, \Sigma, \Delta, q_0, F)$, das gegeben ist durch

- eine endliche Menge Q von Zuständen
- ein Alphabet Σ
- eine Übergangsrelation $\Delta \subseteq Q \times \Sigma^* \times Q$
- einen Anfangszustand (Startzustand) $q_0 \in Q$
- eine Menge $F \subseteq Q$ von Endzuständen

Die vom Automaten akzeptierte Sprache

Definition

Ein nichtdeterministischer endlicher Automat mit Wortübergängen über einem Alphabet Σ **akzeptiert** ein Wort $w \in \Sigma^*$, wenn es Wörter $w_1, \ldots, w_n \in \Sigma^*$ und Zustände $q_0, \ldots, q_n$ gibt, sodass gilt

- q_0 ist der Anfangszustand des Automaten
- $(q_{i-1}, w_i, q_i) \in \Delta$ für $i = 1, \ldots, n$
- $q_n \in F$
- $w = w_1 \ldots w_n$.

Definition

Die von einem nichtdeterministischen endlichen Automaten mit Wortübergängen **akzeptierte Sprache** ist die Menge der von ihm akzeptierten Wörter.

Konfigurationen

Definition

Eine Konfiguration eines nichtdeterministischen endlichen Automaten mit Wortübergängen $(Q, \Sigma, \Delta, q_0, F)$ ist ein Paar (q, w), wobei $q \in Q$ und $w \in \Sigma^*$ ist.

Gibt den momentanen Zustand und den noch zu lesenden Teil des Eingabewortes an.

Berechnungen

Definition

Ein nichtdeterministischer endlicher Automat mit Wortübergängen $(Q, \Sigma, \Delta, q_0, F)$ führt eine Konfiguration (q, w) in eine Konfiguration (q', w') über, wenn es ein Wort $v \in \Sigma^*$ gibt mit

$$w = vw' \wedge (q, v, q') \in \Delta.$$

Berechnungen

Definition

Eine Berechnung durch einen nichtdeterministischen endlichen Automaten mit Wortübergängen $(Q, \Sigma, \Delta, q_0, F)$ ist eine endliche Folge $(k_0, \ldots, k_n)$ von Konfigurationen derart, dass gilt:

- Der Automat führt die Konfiguration k_{i-1} in die Konfiguration k_i über für $i = 1, \ldots, n$.
- k_0 hat die Form (q_0, w).
- k_n hat die Form (q, ε).

Die Berechnung heißt akzeptierend, wenn $q \in F$ ist.

Automaten mit epsilon-Übergängen

Definition

Ein nichtdeterministischer endlicher Automat mit Wortübergängen $(Q, \Sigma, \Delta, q_0, F)$ heißt nichtdeterministischer endlicher Automat mit ε-Übergängen, wenn für alle $(q, v, q') \in \Delta$ gilt: $|v| \leq 1$.

Unterabschnitt 4

Simulation eines Automaten mit Wortübergängen durch einen deterministischen Automaten

Simulation von Wortübergängen durch Zeichenübergänge

Ein Übergang $(q, w, q') \in \Delta$ mit $|w| > 1$ kann unter Verwendung von Zwischenzuständen durch mehrere Zeichenübergänge ersetzt werden.

Beispiel

$(q, abb, q') \in \Delta$

$$q \xrightarrow{abb} q'$$

wird ersetzt durch

$$q \xrightarrow{a} q_1 \xrightarrow{b} q_2 \xrightarrow{b} q'$$

Elimination von epsilon-Übergängen

Epsilon-Übergänge können eliminiert werden, indem jede Folge von ε-Übergängen gefolgt von einem Zeichenübergang ersetzt wird durch einen Zeichenübergang.

Beispiel

$$q \xrightarrow{\varepsilon} q_1 \xrightarrow{\varepsilon} q_2 \xrightarrow{a} q'$$

wird ersetzt durch

$$q \xrightarrow{a} q'$$

Die Menge der Endzustände muss dann noch angepasst werden.

Konstruktion eines deterministischen Automaten

Der letzte Schritt ist die Simulation des so erhaltenen nichtdeterministischen endlichen Automaten durch einen deterministischen, die wir bereits kennen.

Abschnitt 6

Äquivalenz von regulären Ausdrücken und endlichen Automaten

Unterabschnitt 1

Jede von einem endlichen Automaten akzeptierten Sprache ist regulär

Jede von einem endlichen Automaten akzeptierten Sprache ist regulär

Satz

- *Sei L die von einem deterministischen oder nichtdeterministischen endlichen Automaten akzeptierte Sprache.*
- *Dann ist L regulär.*

Beweis

- Für $p, r \in Q$ und $J \subseteq Q$ sei $L[p, J, r]$ die Menge der Wörter, die im endlichen Automaten von p aus durch die Menge J nach r führen.
- Dann ist

$$L = \bigcup_{q \in F} L[q_0, Q, q]$$

Jede von einem endlichen Automaten akzeptierten Sprache ist regulär

Beweis (Fortsetzung)

- $L[p, \emptyset, r]$ ist endlich und daher regulär.
- Es gilt

$$L[p, J \cup \{q\}, r] = L[p, J, r] \cup L[p, J, q]L[q, J, q]^* L[q, J, r].$$

- Durch vollständige Induktion nach der Anzahl der Elemente von J folgt, dass $L[p, J, r]$ regulär ist.
- Also ist die Sprache L regulär. $\square$

Unterabschnitt 2

Jede reguläre Sprachen wird von einem endlichen Automaten akzeptiert

Von endlichen Automaten akzeptierte Sprachen

Definition

- Sei Σ ein Alphabet.
- Dann bezeichnen wir mit $\mathbb{L}_{\mathrm{endA}}(\Sigma)$ die Menge der von endlichen Automaten akzeptierten Sprachen über Σ.

Von endlichen Automaten akzeptierte Sprachen

Lemma

Sei Σ ein Alphabet. Dann gilt:

- $\emptyset \in \mathbb{L}_{\mathrm{endA}}(\Sigma)$.
- $\{\varepsilon\} \in \mathbb{L}_{\mathrm{endA}}(\Sigma)$.
- $\{\iota(a)\} \in \mathbb{L}_{\mathrm{endA}}(\Sigma)$ *für alle* $a \in \Sigma$.
- *Wenn* $L, L' \in \mathbb{L}_{\mathrm{endA}}(\Sigma)$, *dann* $L \cup L' \in \mathbb{L}_{\mathrm{endA}}(\Sigma)$.
- *Wenn* $L, L' \in \mathbb{L}_{\mathrm{endA}}(\Sigma)$, *dann* $LL' \in \mathbb{L}_{\mathrm{endA}}(\Sigma)$.
- *Wenn* $L \in \mathbb{L}_{\mathrm{endA}}(\Sigma)$, *dann* $L^* \in \mathbb{L}_{\mathrm{endA}}(\Sigma)$.
- *Wenn* $L \in \mathbb{L}_{\mathrm{endA}}(\Sigma)$, *dann* $\overline{L} \in \mathbb{L}_{\mathrm{endA}}(\Sigma)$.

Beweisidee

Setze die endlichen Automaten für L und L' mithilfe von ε-Übergängen geeignet zusammen!

Jede reguläre Sprachen wird von einem endlichen Automaten akzeptiert

Folgerung (1)

$\mathbb{L}_{\mathrm{endA}}(\Sigma)$ *ist abgeschlossen nicht nur gegenüber Konkatenation, Mengenvereinigung, Kleene-Stern-Operation und Komplementbildung, sondern auch gegenüber Bildung von Mengendurchschnitt und Mengendifferenz.*

Folgerung (2)

Jede reguläre Sprache über Σ *ist in* $\mathbb{L}_{\mathrm{endA}}(\Sigma)$.

Reguläre Sprachen und endliche Automaten

Satz

Sei L eine Sprache. Dann sind die folgenden Aussagen zueinander äquivalent:

- *L ist regulär.*
- *L wird von einem deterministischen endlichen Automaten akzeptiert.*
- *L wird von einem nichtdeterministischen endlichen Automaten akzeptiert.*
- *L wird von einem nichtdeterministischen endlichen Automaten mit ε-Übergängen akzeptiert.*
- *L wird von einem nichtdeterministischen endlichen Automaten mit Wort-Übergängen akzeptiert.*