Zusammenfassung zur Vorlesung Basismodul Computerlinguistik

Sitzung 1

21.10.2021

Inhaltsverzeichnis

Alphabet und Wort (I)

Alphabet

Eine endliche Menge Σ von Zeichen (Symbolen).
Beispiel: $\Sigma = \{a, b\}$ **Frage**: Kann $\mathbb{N}$ als ein Alphabet betrachtet werden?

Wort

Ein Wort über einem Alphabet Σ ist eine endliche geordnete Folge von Symbolen aus Σ.
Beispiel: $w = aaba$ ist ein Wort über dem Alphabet $\Sigma = \{a, b\}$.

- **Länge eines Wortes**: $|w| = 4$
- **Leeres Wort**: ϵ, $|\epsilon| = 0$

- Σ^n: die Menge aller Wörter der Länge n über Σ.
- Σ^*: die Menge aller Wörter über Σ mit beliebiger Länge (inklusiv des leeren Wortes ϵ).
- Σ^+: die Menge aller nicht-leeren Wörter über Σ.

Alphabet und Wort (II)

Konkatenation der Wörter

Die *Konkatenation* $\circ$: Wörter werden hintereinander geschrieben.

- Formale Definition
  $\circ : \Sigma^* \times \Sigma^* \rightarrow \Sigma^*$ ist für zwei Wörter über Σ $|u| = k$ und $|v| = l$ wie folgt definiert:

$$u \circ v(i) = \begin{cases} u(i) & \text{wenn } 1 \leq i \leq k \\ v(i-k) & \text{wenn } k < i \leq k+l \end{cases}$$

- Beispiel
  $u = abc, v = bcd$

$\rightarrow u \circ v = abcbcd$

Sprache*

***Achtung**: Hier bezeichnet Sprache die formale Sprache statt der natürlichen Sprache und hat die andere Bedeutung als in der Linguistik.

Sprache

Die Sprache L über Σ^* ist eine Teilmenge von Σ^*.

- *Konkatenation der Sprache* $L_1 \circ L_2 := \{w_1 w_2 | w_1 \in L_1, w_2 \in L_2\}$
- Die leere Sprache $L := \{\}$ oder $\emptyset$
- Induktive Definition von n Faktoren der Sprache

$$L^0 := \{\epsilon\}$$

$$L^{n+1} := L^n \circ L$$

- *Kleene-Stern* oder *Konkatenationsabschluss* einer Sprache:

$$L^* := \cup_{n \in \mathbb{N}} L^n$$

  Alle Konkatenationen von 0, 1,... Wörtern aus L.

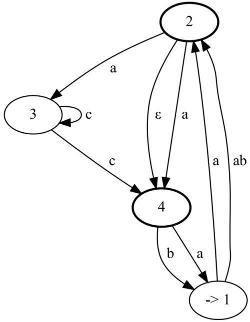
Grundbegriffe

Endlicher Automat (EA)*

Ein EA ist ein 5-Tupel $A = (\Sigma, Q, I, F, \Delta)$ mit:

- σ: einem endlichen Alphabet
- Q: einer endlichen und nicht-leeren Zustandsmenge
- $I \subseteq Q$: einer Menge von Initialzuständen
- $F \subseteq Q$: einer Menge von Finalzuständen
- $\Delta \subseteq Q \times \Sigma^* \times Q$: einer Übergangsrelation (Hier bezeichnet das Symbol $\times$ den Kartesischen Produkt.)

* wird auch Nicht-Deterministischer Endlicher Automat (NDEA) genannt.



Beispiel von EA

- $Q = \{1, 2, 3, 4\}$
- $F = \{2, 4\}$
- $I = \{1\}$
- $\Sigma = \{a, b, c\}$

Sprache eines Automates $L(A)$

Um eine Sprache des Automates zu definieren, müssen wir zunächst zwei andere Begriffe kennen, nämlich **Pfad** und **Label**.

Pfad und Label

- Ein Pfad ist eine Folge von n Übergängen der Form

$$p_0 \xrightarrow{w_1} p_1, p_1 \xrightarrow{w_2} p_2, \cdots, p_{n-1} \xrightarrow{w_n} p_n$$

- Das Label des Pfads von p_0 nach p_n ist $w1 \circ w2 \circ \cdots \circ w_n$.

Von A akzeptierte Sprache

- Die Sprache $L(A) =$ Menge aller Labels von erfolgreichen Pfaden.

Sprache eines Zustands p $L(p)$

Sei $p \in Q$, $L(p) =$
Menge aller Labels von Pfaden, die von p in Finalzuständen führen.

Induktive Definition der Sprache eines Zustands p

$$L(P) = \cup_{(p,w,q)\in\Delta} w \circ L(q) \cup \begin{cases} \{\epsilon\} & \text{falls } p \in F \\ \emptyset & \text{sonst} \end{cases}$$

Äquivalenz

Verallgemeinte Übergangsrelation Δ^* - Induktive Definition

- Für jedes $p \in Q$: $(p, \epsilon, p) \in \Delta^*$
- Für $(p, w, r) \in \Delta^*, (r, v, q) \in \Delta$, so auch $(p, wv, q) \in \Delta^*$

Äquivalente Automaten

Es gilt $L(A) = \{w \in \Sigma^* | \exists i \in, f \in F, (i, w, f) \in \Delta^*\}$,

$$L(A_1) = L(A_2) \iff A_1 \text{ und } A_2 \text{sind äquivalente Automaten}$$

Äquivalente Zustände

Sei $p \in Q$, $L(p) = \{w \in \Sigma^* | \exists f \in F, (p, w, f) \in \Delta^*\}$,

$$\text{Seien } p, q \in Q, L(p) = L(q) \iff p \text{ und } q \text{sind äquivalente Zustände}$$

.