Grundlagen der theoretischen Informatik

Kurt Sieber

Fachbereich Mathematik/Theoretische Informatik
Universität Siegen

Inhalt der Vorlesung

Teil I: Automaten und formale Sprachen

Zentrale Fragestellungen

- Wie definiert man (formal) die Syntax einer Sprache (z.B. einer Programmiersprache)?
- Wie überprüft man, ob ein Wort (d.h. eine Zeichenreihe) zur Sprache gehört?

Teil II: Berechenbarkeit und Komplexität

Zentrale Fragestellungen

- Welche Probleme (Aufgabenstellungen) sind prinzipiell mit einem einem Computer lösbar bzw. nicht lösbar?
- Welche Probleme sind mit vernünftigem Aufwand (an Speicherplatz und Laufzeit) lösbar?

Grundlagen für beide Teile: Zeichen, Wörter und Sprachen

Definition 0.1 *Ein Alphabet oder Zeichenvorrat ist eine nichtleere endliche Menge. Die Elemente eines Alphabets nennen wir Zeichen oder Symbole.*

Konvention: Alphabete werden mit $\Sigma, \Sigma_1, \Sigma', \ldots$ bezeichnet.

Zeichen, Wörter, Sprachen

Beispiele

- $\Sigma_1 = \{1\}$
- $\Sigma_2 = \{0, 1\}$
- $\Sigma_3 = \{0, \ldots, 9\}$
- $\Sigma_4 = \{a, \ldots, z, A, \ldots, Z\}$
- $\Sigma_5 =$ Menge aller ASCII-Zeichen
- $\Sigma_6 = \{\textit{begin}, \textit{end}, \textit{if}, \textit{then}, \textit{else}, ;, :=, +, *, \ldots\}$

Zeichen, Wörter, Sprachen

Definition 0.2 *Ein Wort über dem Alphabet Σ ist eine endliche Folge $w = (a_1, \ldots, a_n)$, wobei $n \geq 0$ und $a_i \in \Sigma$ für $i = 1, \ldots, n$. n heißt Länge des Wortes w, a_i heißt i-tes Zeichen von w. Das Wort $()$ heißt leeres Wort.*

Kurzschreibweise

- $a_1 \ldots a_n$ statt $(a_1, \ldots, a_n)$
  Vorsicht: Zeichen a und Wort a nicht mehr unterscheidbar!
- ε statt $()$

Weitere Schreibweisen

- $|w| =$ Länge von w
- $\Sigma^* =$ Menge aller Wörter über Σ
- $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\} =$ Menge aller nichtleeren Wörter über Σ

Beispiele

- $\{1\}^* = \{\varepsilon, 1, 11, 111, \ldots\}$
- $\{0,1\}^* = \{\varepsilon, 0, 1, 00, 01, 10, 11, \ldots\}$
- $\varepsilon, 0, 1000, 0001 \in \{0, \ldots, 9\}^*$
- Der Inhalt einer Datei ist ein Wort über dem ASCII-Alphabet (denn Leerzeichen, Zeilenwechsel etc. sind Zeichen dieses Alphabets).
- Ein Programm (in irgendeiner Programmiersprache) ist ebenfalls ein Wort über dem ASCII-Alphabet.
- Der Inhalt eines Buches ist ein Wort (über dem Alphabet der Sprache, in der es geschrieben ist).

Alternative Definition für Σ^* und Σ^+

- $$\begin{aligned} \Sigma^n &= \{(a_1, \ldots, a_n) \mid a_i \in \Sigma \text{ für } i = 1, \ldots, n\} \\ &= \{a_1 \ldots a_n \mid a_i \in \Sigma \text{ für } i = 1, \ldots, n\} \\ &= \text{Menge aller Wörter der Länge } n \text{ über } \Sigma \end{aligned}$$

- Speziell: $$\begin{aligned} \Sigma^0 &= \{()\} \\ &= \{\varepsilon\} \end{aligned}$$

- $$\begin{aligned} \Sigma^* &= \textstyle\bigcup_{n \geq 0} \Sigma^n \\ &= \text{Menge aller Wörter über } \Sigma \\ \Sigma^+ &= \textstyle\bigcup_{n \geq 1} \Sigma^n \\ &= \Sigma^* \setminus \{\varepsilon\} \end{aligned}$$

Definition 0.3 *Seien* $v = a_1 \dots a_m$ *und* $w = b_1 \dots b_n$ *Wörter über* Σ. *Dann ist die Konkatenation* $v \circ w$ *von* v *und* w *definiert durch*

$$v \circ w \;=\; a_1 \dots a_m b_1 \dots b_n$$

Kurzschreibweise: vw statt $v \circ w$

Eigenschaften der Konkatenation

- Assoziativität: $(uv)w = u(vw)$ für alle $u, v, w \in \Sigma^*$
- Neutrales Element: $w\varepsilon = \varepsilon w = w$ für alle $w \in \Sigma^*$
- Also: $(\Sigma^*, \circ)$ ist ein Monoid mit neutralem Element ε
(vgl. natürliche Zahlen mit Multiplikation)

Definition 0.4 *Sei $w \in \Sigma^*$ und $n \in \mathbb{N}$. Dann ist die* *n-te Potenz* w^n *von w definiert durch $w^n = \underbrace{w \ldots w}_{n}$ (speziell: $a^n = \underbrace{a \ldots a}_{n}$ falls $a \in \Sigma$).*

Besser: Definition von w^n durch Induktion über n

- $w^0 = \varepsilon$
- $w^n = ww^{n-1}$ falls $n > 0$

Oder:

- $w^0 = \varepsilon$
- $w^n = w^{n-1}w$ falls $n > 0$

Beispiele

- $(bla)^2 = blabla$
- $(bla\ )^3 = bla\ bla\ bla$

Definition 0.5 *Für $w = a_1 \ldots a_n \in \Sigma^*$ sei $w^R \in \Sigma^*$ das Wort, das durch "Spiegelung" aus w entsteht, d.h.*

$$w^R = a_n \ldots a_1$$

(R steht für engl. reverse)

Beispiel

$(01001)^R = 10010$

Alternative

Definition von w^R durch Induktion über $|w|$
(s. Übung!)

Definition 0.6 *Sei $x \in \Sigma^*$.*

- u *heißt* *Anfangswort* *oder* *Präfix* *von x, falls ein $v \in \Sigma^*$ existiert mit $x = uv$.*
- v *heißt* *Endwort* *oder* *Suffix* *von x, falls ein $u \in \Sigma^*$ existiert mit $x = uv$.*
- v *heißt* *Teilwort* *von x, falls $u, w \in \Sigma^*$ existieren mit $x = uvw$.*

Beispiel

- Anfangsworte von 010: $\varepsilon, 0, 01, 010$
- Endworte von 010: $\varepsilon, 0, 10, 010$
- Teilworte von 010: $\varepsilon, 0, 1, 01, 10, 010$

Die Beziehungen 'Anfangswort', 'Endwort' und 'Teilwort' sind zweistellige Relationen auf Σ^*. Jede dieser Relationen ist

- reflexiv: Jedes $w \in \Sigma^*$ steht in Relation zu sich selbst.
- transitiv: Wenn u in Relation zu v und v in Relation zu w steht, dann steht auch u in Relation zu w.
- antisymmetrisch: Wenn u in Relation zu v steht und v in Relation zu u, dann ist $v = u$.

Also: Jede der Relationen definiert eine partielle Ordnung auf Σ^*.

Definition 0.7 *Eine Sprache (oder formale Sprache) über dem Alphabet Σ ist eine (beliebige) Teilmenge von Σ^*.*

Konvention:

Sprachen bezeichnen wir meist mit $L, L_1, L', \ldots$ (engl. language).

Alternative Formulierung

Eine Sprache über Σ ist ein Element der Potenzmenge $\wp(\Sigma^*)$.

Zeichen, Wörter, Sprachen

Beispiele

- $L_1 = \emptyset$
- $L_2 = \Sigma^*$
- $L_3 = \{\varepsilon\}$
- $L_4 = \{1^n \mid n \text{ ist eine Primzahl}\}$
- $L_5 =$ Menge aller Binärdarstellungen natürlicher Zahlen
- $L_6 =$ Menge aller Dezimaldarstellungen natürlicher Zahlen
- $L_7 =$ Menge aller Java-Programme
- $L_8 =$ Menge aller Sätze der deutschen Sprache
- $L_9 = \{0^n 1^n \mid n \geq 0\}$
- $L_{10} = \{vw \mid v \in \{0\}^*, w \in \{1\}^*\}$

Zeichen, Wörter, Sprachen

Definition 0.8 *Seien $L, L_1, L_2 \subseteq \Sigma^*$. Dann sei*

- $L_1 \cup L_2$ *die Vereinigung von L_1 und L_2.*
- $L_1 \cap L_2$ *der Durchschnitt von L_1 und L_2.*
- $L_1 \setminus L_2 = \{w \in L_1 \mid w \notin L_2\}$ *die Mengendifferenz von L_1 und L_2.*
- $\bar{L} = \Sigma^* \setminus L$ *das Komplement von L.*
- $L_1 \circ L_2 = \{vw \mid v \in L_1, w \in L_2\}$ *die Konkatenation von L_1 und L_2.*

Kurzschreibweise: L_1L_2 statt $L_1 \circ L_2$

Eigenschaften der Konkatenation

- Assoziativität: $(L_1L_2)L_3 = L_1(L_2L_3)$ für alle $L_1, L_2, L_3 \subseteq \Sigma^*$
- Neutrales Element: $L\{\varepsilon\} = \{\varepsilon\}L = L$ für alle $L \subseteq \Sigma^*$
- Also: $(\wp(\Sigma^*), \circ)$ ist ein Monoid mit neutralem Element $\{\varepsilon\}$.

Definition 0.9 *Seien $L \subseteq \Sigma^*$ und $n \in \mathbb{N}$. Dann ist die n-te Potenz L^n von L definiert durch*

$$\begin{aligned} L^n &= \underbrace{L \circ \ldots \circ L}_{n} \\ &= \underbrace{L \ldots L}_{n} \\ &= \{w_1 \ldots w_n \mid w_i \in L \text{ für } i = 1, \ldots, n\} \end{aligned}$$

Besser: Definition von L^n durch Induktion über n

- $L^0 = \{\varepsilon\}$
- $L^n = LL^{n-1}$ falls $n > 0$

Oder:

- $L^0 = \{\varepsilon\}$
- $L^n = L^{n-1}L$ falls $n > 0$

Definition 0.10 *Für $L \subseteq \Sigma^*$ sei*

- $L^R = \{w^R \mid w \in L\}$ *die "Spiegelung" von* L
- $$\begin{aligned} L^* &= \bigcup_{n\geq 0} L^n \\ &= \{w_1 \ldots w_n \mid n \geq 0, w_i \in L \text{ für } i = 1, \ldots, n\} \end{aligned}$$
  der Kleene-Abschluss von L.
- *Speziell:*
  $$\begin{aligned} \{w\}^* &= \{w^n \mid n \geq 0\} \text{ falls } w \in \Sigma^* \\ \{a\}^* &= \{a^n \mid n \geq 0\} \text{ falls } a \in \Sigma \end{aligned}$$
- $$\begin{aligned} L^+ &= \bigcup_{n\geq 1} L^n \\ &= \{w_1 \ldots w_n \mid n \geq 1, w_i \in L \text{ für } i = 1, \ldots, n\} \\ &= LL^* \end{aligned}$$

Folgerung: $L^* = L^+ \cup L^0 = L^+ \cup \{\varepsilon\}$