Kapitel 1
Formale Sprachen und Automaten

1.1 Sprachen und Grammatiken

Grundbegriffe

Alphabet: endliche Menge Σ; Elemente heißen (in diesem Kontext) ***Zeichen*** (***Symbole***).

- Σ^*: Menge aller ***Zeichenreihen*** (***Wörter***, ***Strings***) über Σ.
- ε: ***leere Zeichenreihe***.
- $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$.

Länge $|w|$ einer Zeichenreihe w: Anzahl der in w auftretenden Zeichen ($|\varepsilon| = 0$).

- $|w|_a$ ($a \in \Sigma$): Anzahl der Vorkommen von a's in w.

Konkatenation von Zeichenreihen: $\circ : \Sigma^* \times \Sigma^* \to \Sigma^*$.

- $u, v \in \Sigma^*$: u ***Teilzeichenreihe*** (***Teilwort***, ***Teilstring***) von v: es gibt $x, y \in \Sigma^*$ mit $v = x \circ u \circ y$
  ($x = \varepsilon$: u ***Präfix*** von v; $y = \varepsilon$: u ***Suffix*** von v).

Spiegelung von Zeichenreihen: $R : \Sigma^* \to \Sigma^*,\ w \mapsto w^R$,

definiert durch $\quad \varepsilon^R = \varepsilon,$
$$(a \circ u)^R = u^R \circ a \quad \text{für } a \in \Sigma, u \in \Sigma^*.$$

Es gilt:

1. $|u \circ v| = |u| + |v|$.
2. $w \in \Sigma^*$: ε und w sind Präfixe und Suffixe von w.

Schreibweisen

- uv statt $u \circ v$, uvw statt $((uv)w)$, ...
- Zeichenreihen: $a_1 a_2 \dots a_n$, ua, $ua_1 \dots a_n v$, ...
  ($a_i \in \Sigma$; $u, v \in \Sigma^*$).
- w^n für $\underbrace{ww \dots w}_{n-\text{mal}}$ ($w^0 = \varepsilon$) für $w \in \Sigma$ oder $w \in \Sigma^*$.

Definition. Eine (***formale***) ***Sprache*** über einem Alphabet Σ ist eine Teilmenge von Σ^*.

Operationen auf Sprachen

Durchschnitt $\cap$, ***Vereinigung*** $\cup$, ***Komplement*** $^-$ ($\bar{L} = \Sigma^* \setminus L$).

Sprachprodukt (Konkatenation): $L \circ M = \{u \circ v \mid u \in L, v \in M\}$.

Sprachpotenz: $L^0 = \{\varepsilon\}$, $L^{n+1} = L \circ L^n \quad (n \in \mathbb{N}_0)$.

Kleene-Stern: $L^* = \bigcup_{n \geq 0} L^n \qquad$ (außerdem: $L^+ = \bigcup_{n > 0} L^n$).

Spiegelung: $L^R = \{w^R \mid w \in L\}$.

Beispiel einer Syntaxdefinition

⟨Satz⟩ $\to$ ⟨Subjekt⟩ ⟨intransitives Prädikat⟩
⟨Subjekt⟩ $\to$ ⟨Artikel⟩ ⟨Attribut⟩ ⟨Substantiv⟩
⟨Artikel⟩ $\to$ der | die | das
⟨Attribut⟩ $\to$ ε | hungrige | große
⟨Substantiv⟩ $\to$ Katze | Mensch | Tier
⟨intransitives Prädikat⟩ $\to$ ⟨Verb⟩ ⟨Adverb⟩
⟨Verb⟩ $\to$ schläft | singt | läuft
⟨Adverb⟩ $\to$ laut | leise | schnell

Definition. Eine ***Grammatik*** (***Phrasenstrukturgrammatik***) $G = (V, \Sigma, P, S)$ ist gegeben durch:

- V und Σ sind zwei Alphabete mit $V \cap \Sigma = \emptyset$;
  die Elemente von V heißen (***syntaktische***) ***Variablen*** (***Nichtterminalzeichen***), die von Σ ***Terminalzeichen***.
- $P \subseteq (V \cup \Sigma)^+ \times (V \cup \Sigma)^*$ ist eine endliche Menge von ***Regeln*** (***Produktionen, Produktionsregeln***);
  Schreibweise (meist) $u \to_G v$ bzw. $u \to v$ statt $(u, v) \in P$.
- $S \in V \qquad$ (***Startvariable, Axiom***).

Definition. Sei $G = (V, \Sigma, P, S)$ eine Grammatik. Die Relation $\Rightarrow_G \subseteq (V \cup \Sigma)^* \times (V \cup \Sigma)^*$ ist gegeben durch:

$$u \Rightarrow_G v \iff \text{es gibt } x, y, y', z \in (V \cup \Sigma)^*, \text{ so dass } u = xyz, v = xy'z \text{ und } y \to_G y'.$$

Sei $\Rightarrow_G^*$ die reflexive, transitive Hülle von $\Rightarrow_G$. Die Sprache $\mathcal{L}(G) = \{w \in \Sigma^* \mid S \Rightarrow_G^* w\}$ heißt die ***von G erzeugte Sprache***. Zwei Grammatiken G_1, G_2 heißen ***äquivalent***, wenn $\mathcal{L}(G_1) = \mathcal{L}(G_2)$ gilt.