Formale Sprachen und Komplexitätstheorie

WS 2019/20

Robert Elsässer

4. Formale Sprachen

Grammatiken

- dienen zur formalen Beschreibung von Sprachen
- werden benutzt zur Beschreibung von Programmiersprachen, Anfragesprachen bei Datenbanken, usw.
- sollen einfach sein, um eine effiziente Analyse zu erlauben
- sollen mächtig genug sein, um aufwändige Konstrukte wie z.B. geschachtelte Schleifen beschreiben zu können.
- Wortprobleme bei eingeschränkten Grammatiken führen zu speziellen Rechenmodellen, die genau diese Wortprobleme lösen.

4. Formale Sprachen

Grammatiken

- Rechenmodelle sind endliche Automaten für reguläre Grammatiken und Kellerautomaten für kontextfreie Grammatiken.
- Varianten von kontextfreien Grammatiken werden zur Beschreibung von Programmiersprachen eingesetzt.
- Reguläre Grammatiken und endliche Automaten werden bei Kontrollsystemen und in der lexikographischen Analyse eingesetzt.
- Allgemeine Grammatiken liefern alternative Beschreibungen von rekursiv aufzählbaren Sprachen.

Grammatiken

Definition

Eine Grammatik (vom Typ Chomsky-0) ist ein 4-Tupel (V, Σ, P, S), für den gilt:

- V ist ein endliches Alphabet von Variablen
- Σ ist ein endliches Alphabet von Terminalen
- S ist das Startsymbol
- P ist eine endliche Menge von Produktionen oder Ersatzregeln,
  d.h. P ist eine Teilmenge von $\left((V \cup \Sigma)^+ \setminus \Sigma^*\right) \times (V \cup \Sigma)^*$

Grammatiken

- w' ist aus w direkt ableitbar, wenn es eine Ersetzungsregel $u \to v$ und α, β in $(V \cup \Sigma)^*$ gibt, so dass $w = \alpha u \beta$ und $w' = \alpha v \beta$, geschrieben $w \to w'$.

- w' ist aus w ableitbar, falls w' durch endlich viele Ableitungsschritte aus w erhalten werden kann, geschrieben $w \xrightarrow{*} w'$

Äquivalent: es gibt $w_0 = w, w_1, \dots, w_{n-1}, w_n = w'$ mit $w_{i-1} \to w_i$

Grammatiken

Definition

Sei $G = (V, \Sigma, P, S)$ eine Grammatik. Dann ist

$$L(G) \coloneqq \{w \text{ aus } \Sigma^* \mid S \xrightarrow{*} w\}$$

die von G erzeugte Sprache.

Grammatiken

In einer Linksableitung wird in jedem Schritt die am weitesten links stehende Variable im nächsten Schritt ersetzt.

Ableitungen sind in der Regel nicht eindeutig.

Grammatiken

Satz

Eine Sprache L ist genau dann rekursiv aufzählbar, wenn es eine Grammatik G vom Typ Chomsky-0 gibt mit $L(G) = L$.

Eingeschränkte Grammatiken

Definition

- Eine Grammatik heißt kontextsensitiv oder vom Typ Chomsky-1, falls für jede Regel $u \to v$ gilt: $|u| \leq |v|$.

- Eine Grammatik heißt kontextfrei oder vom Typ Chomsky-2, falls für jede Regel $u \to v$ gilt: $u \in V$.

- Eine Regel heißt regulär oder vom Typ Chomsky-3, falls alle Regeln der Art $u \to v$ mit $u \in V$ und:
  - $v = \varepsilon$
  - $v = a$, $a \in \Sigma$ oder
  - $v = aw$ mit $a \in \Sigma$ und $w \in V$

  sind.

Eingeschränkte Grammatiken

Definition

- Eine Grammatik heißt kontextsensitiv oder vom Typ Chomsky-1, falls für jede Regel $u \to v$ gilt: $|u| \leq |v|$.

Ausnahme: Bei kontextsensitiven Grammatiken wird die Regel $S \to \varepsilon$ zugelassen.
Es muss dann allerdings für alle Regeln $u \to v$ gelten, dass S in v nicht vorkommt.

4. Formale Sprachen

Definition

Eine Sprache L heißt kontextsensitiv, kontextfrei oder regulär, wenn es eine kontextsensitive, kontextfreie oder reguläre Grammatik G gibt mit $L(G) = L$.

Satz

Jede kontextsensitive Sprache ist entscheidbar.