Grundzüge der Theoretischen Informatik

26.1.2022

Markus Bläser

Universität des Saarlandes

Noam Chomsky

Natursprachliche Sätze strukturieren

< Subjekt >< Prädikat > < Objekt >

Kapitel 28: Grammatiken

Grammatiken

(u,w) ∈ P
„u kann durch w ersetzt werden"

Definition (28.1)

Eine Grammatik G ist ein 4-Tupel (V, Σ, P, S):

1. V ist eine endliche Menge, die *Variablen* oder *Nichterminale*.
2. Σ ist eine endliche Menge, die *Terminale*.
   Es gilt $V \cap \Sigma = \emptyset$.
   nichtleere Wörter
3. P ist eine endliche Teilmenge von $(V \cup \Sigma)^+ \times (V \cup \Sigma)^*$, die *Produktionen*.
4. $S \in V$ ist die *Startvariable*.

Kovention 28.2

Für $(u, v) \in P$ schreiben wir auch $u \to v$ statt (u, v).

Ableitungen und erzeugte Sprache

$abbab \Rightarrow abbbab$

$(ba, bba) \in P$

„kann in einem Schritt abgeleitet werden“

Definition (28.3)

1. Eine Grammatik $G = (V, \Sigma, P, S)$ definiert eine binäre Relation $\Rightarrow_G$ auf $(V \cup \Sigma)^*$ wie folgt: $u \Rightarrow_G v$ falls wir $u = xyz$ und $v = xy'z$ schreiben können mit $y \to y' \in P$. Wir sagen v ist *ableitbar* aus u (in einem Schritt). v ist ableitbar aus u falls $u \Rightarrow_G^* v$.
   $u \Rightarrow u_1 \Rightarrow u_2 \Rightarrow \ldots \Rightarrow v$
2. $u \in (V \cup \Sigma)^*$ heißt *Satzform*, falls $S \Rightarrow_G^* u$.
3. Eine Folge von Satzformen $w_0, \ldots, w_t \in (V \cup \Sigma)^*$ mit $w_0 = S$, $w_\tau \Rightarrow_G w_{\tau+1}$ für $0 \leq \tau < t$ und $w_t = u$ heißt *Ableitung* von u.
4. Die von G erzeugte Sprache ist $L(G) = \{u \in \Sigma^* \mid S \Rightarrow_G^* u\}$.

Beispiel 28.4

Sei $G_1 = (\{S\}, \{0, 1\}, P_1, S)$ mit P_1 gegeben durch

$$S \to \varepsilon$$
$$S \to 0S1$$

$$\begin{array}{ccccccccc}
S & \Rightarrow & 0S1 & \Rightarrow & 00S11 & \Rightarrow & 000S111 & & \cdots \\
\Downarrow & & \Downarrow & & \Downarrow & & \vdots & & \\
\varepsilon & & 01 & & 0011 & & & & \cdots
\end{array}$$

$$L(G_1) = \{ 0^n 1^n \mid n \in \mathbb{N} \}$$

Beispiel 28.6

$u \to v_1 ,\ u \to v_2 .\ u \to v_3 \ \ldots$

$u \to v_1 | v_2 | v_3 | \ldots$

$P_2 \subseteq (V \cup \Sigma)^+ \times (V \cup \Sigma)^*$

Sei $G_2 = (\{W, V, N, N^+, Z, Z^+\}, \{a, b, \ldots, z, 0, 1, \ldots, 9, :, =, \neq, ;, +, -, [,]\}, P_2, W)$ mit P_2 gegeben durch

$$
\begin{aligned}
W &\longrightarrow V{:=}V{+}V \mid V{:=}V{-}V \mid V{:=}N^+ \mid \\
&\quad\ \texttt{while}\ V{\neq}\,0\ \texttt{do}\ W\ \texttt{od} \mid \\
&\quad\ [W;W] \\
V &\longrightarrow \mathtt{x}N^+ \\
N^+ &\longrightarrow Z \mid Z^+N \\
N &\longrightarrow Z \mid ZN \\
Z &\longrightarrow 0 \mid 1 \mid \cdots \mid 9 \\
Z^+ &\longrightarrow 1 \mid 2 \mid \cdots \mid 9
\end{aligned}
$$

$\cancel{\mathbb{N}^+}$

alle Ziffernfolge ohne führende Nullen. $\Rightarrow$ Kodierung von $\mathbb{N}$

$N \Rightarrow 0 \ldots 5\, N$

$0 \ldots 9$

Beispiel 28.7

$L(G_3) = \{0^n 1^n 2^n \mid n \in \mathbb{N}\}$

Sei $G_3 = (\{S, E, Z\}, \{0, 1, 2\}, P_3, S)$ mit P_3 gegeben durch

$$
\begin{aligned}
S &\rightarrow 0EZ \mid 0SEZ \\
ZE &\rightarrow EZ \\
0E &\rightarrow 01 \\
1E &\rightarrow 11 \\
1Z &\rightarrow 12 \\
2Z &\rightarrow 22
\end{aligned}
$$

$S \Rightarrow 0SEZ \Rightarrow 00SEZEZ$

$S \Rightarrow 0EZ \Rightarrow 01Z \Rightarrow 012$

$0SEZ \Rightarrow 00EZEZ \Rightarrow 001ZEZ \Rightarrow 0012EZ$ ✗

$001ZEZ \Rightarrow 001EZZ \Rightarrow 0011ZZ \Rightarrow 00112Z \Rightarrow 001122$

Die Chomsky-Hierarchie

$u \to v_1$
$u \to v_2$

$a \to AB$

$A \to \varepsilon$
$B \to \varepsilon$

Definition (28.8)

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

1. Jede Grammatik G ist eine *Typ-0-Grammatik*. $\hat{=}$ RE
2. G ist eine *Typ-1-Grammatik*, falls $|u| \leq |v|$ für jedes $u \to v \in P$. „nicht verkürzend" $\hat{=}$ NSPACE(n)
   Die einzige Ausnahme ist $S \to \varepsilon$. Falls $S \to \varepsilon \in P$, dann erscheint S auf keiner rechten Seite.
3. G ist eine *Typ-2-Grammatik* falls sie Typ-1 ist und die linke Seite jeder Produktion aus V ist. kontextfrei
4. G ist eine *Typ-3-Grammatik* falls sie Typ-2 ist und jede rechte Seite einer Produktion $\Sigma V \cup \Sigma$. „rechtslinear" $\hat{=}$ REG

Die Chomsky-Hierarchie (2)

Definition (28.9)

Sei $i \in \{0, 1, 2, 3\}$. $L \subseteq \Sigma^*$ ist eine *Typ-i-Sprache*, falls es eine Typ-i-Grammatik G gibt mit $L = L(G)$.

Definition (28.10)

1. Die Menge aller Typ-2-Sprachen wird mit CFL bezeichnet.
2. Die Menge aller Typ-1-Sprachen wird mit CSL bezeichnet.

Theorem (28.12)

L *ist eine Typ-3-Sprache genau dann, wenn* L *regulär ist.*

Theorem (VI.6)

L *ist eine Typ-0-Sprache genau dann, wenn* $L \in \mathrm{RE}$.

Beweis 28.12 Typ 3 $\Rightarrow$ regulär , $\Leftarrow$ Übung

Sei $G = (V, \Sigma, P, S)$ eine Typ 3 - Grammatik für L.

Erst einmal nehmen wir an, dass $\varepsilon \notin L$

Sei $F \notin V$ $\qquad\qquad \delta : Q \times \Sigma \to \mathcal{P}(Q)$

Sei $M = (V \cup \{F\}, \Sigma, \delta, S, \{F\})$

wobei δ definiert ist durch:

$$\delta(A, \sigma) = \begin{cases} \{B \mid A \to \sigma B\} & \text{falls } A \to \sigma \notin P \\ \{B \mid A \to \sigma B\} \cup \{F\} & \text{sonst} \end{cases}$$

für alle $A \in V$ und $\sigma \in \Sigma$

Sei: $w = w_1 w_2 \dots w_n \in \Sigma^*$

$w \in L(G) \iff S \Rightarrow^* w$

$\iff \exists$ Variablen $V_1, \dots, V_{n-1} \in V$ mit:

$V_1 \to w_2 V_2 \in P$

$S \Rightarrow w_1 V_1 \Rightarrow w_1 w_2 V_2 \Rightarrow \dots$ $V_{n-1} \to w_n$

$\dots \Rightarrow w_1 w_2 \dots w_{n-1} V_{n-1} \Rightarrow w_1 w_2 \dots w_n$

$\iff \exists$ Zustände $V_1, \dots, V_{n-1} \in V$:

$V_1 \in \delta(S, w_1)$, $V_2 \in \delta(V_1, w_2)$, ...

$V_{n-1} \in \delta(V_{n-2}, w_{n-1})$, $\mp \in \delta(V_{n-1}, w_n)$

$\iff \exists$ akz. Berechnung von M auf w

$\iff w \in L(M)$ $\square$

Kapitel 29: Kontextfreie Grammatiken

Ableitungsbäume und Mehrdeutigkeit

$$E \to E * E \mid E + E \mid (E) \mid x$$

Beispiel: $x + x * x$

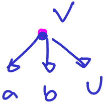
Linksableitung:

$$E \Rightarrow E + E \Rightarrow x + E \Rightarrow x + E * E \Rightarrow x + x * E \Rightarrow x + x * x$$

Rechtsableitung:

$$E \Rightarrow E + E \Rightarrow E + E * E \Rightarrow E + E * x \Rightarrow E + x * x \Rightarrow x + x * x$$

Ableitungsbäume



$V \to abV$

Definition (29.1)

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

1. Ein *Ableitungsbaum* ist ein geordneter Baum mit einer Knotenbeschriftung:
   1.1 Die Wurzel ist mit S beschriftet.
   1.2 Alle Blätter sind mit $V \cup \Sigma$ beschriftet oder mit ε. Im letzten Fall gibt es nur ein Blatt.
   1.3 Alle inneren Knoten sind mit Elementen aus V beschriftet. Falls A eine Knotenbeschriftung ist und $x_1, x_2, \ldots, x_t$ (in dieser Reihenfolge) die Beschriftungen der Kinder, dann ist $A \to x_1 x_2 \ldots x_t \in P$.
2. Das *Blattwort* ist die Konkatenation der Beschriftungen der Blätter.

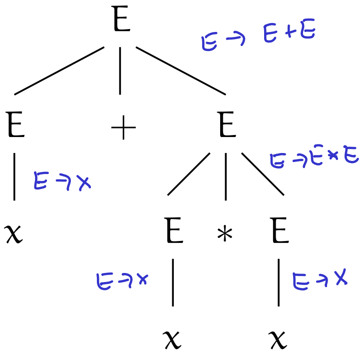
Beispiel

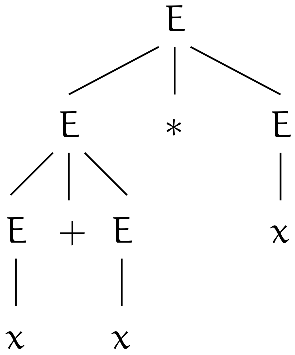
x + x * x

E

↓

E * E





Eindeutigkeit

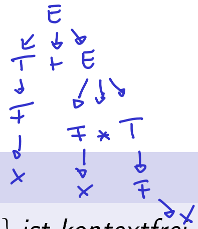
$2+3=3+2$

Definition (29.2)

1. Eine kfG heißt *mehrdeutig*, falls ein Wort zwei oder mehr Ableitungsbäum hat. Sonst heißt sie *eindeutig*.
2. Eine kfS L heißt *eindeutig*, falls es eine eindeutige Grammatik für L gibt. Sonst ist Sie *inhärent mehrdeutig*.

$x + x * x$

$$E \longrightarrow T \mid T + E$$
$$T \longrightarrow F \mid F * T$$
$$F \longrightarrow x \mid (E)$$



$S \to S_1 S_2$

$S \to S_1 \mid S_2$

Theorem (29.3)

$\{0^n 1^n 2^m 3^m \mid n, m \geq 1\} \cup \{0^n 1^m 2^m 3^n \mid n, m \geq 1\}$ *ist kontextfrei und inhärent mehrdeutig.*