Vorlesungsmitschrift zur Vorlesung Theoretische Informatik vom 14.07.2000

Von Matr.-Nr.: 702100

5.1.2. Chomsky-Normalform

kontextfreier Fall: $A \rightarrow w$, $w \in (N \cup T)^*$, $A \subseteq N$

$\Rightarrow$ Chomsky-Normalform: **1**. ε–Regeln vermeiden, soweit möglich
2. Rechte-Seiten vereinfachen
- $|w| \leq 2$
- $|w| = 2$, $w \in N \bullet N$
- $|w| \geq 1$, $w \in T$

Def. E: Eine kontextfreie Grammatik $G = (N, T, P, \sigma)$ heisst Chomsky-Normalform wenn gilt:
- Falls $\varepsilon \in L(G)$, so gibt es die Regeln $\sigma \rightarrow \varepsilon$ und σ kommt auf keiner rechten Seite einer Produktion vor
- Für alle $\alpha \rightarrow \beta \in P$ gilt: $\alpha \in N$, $\beta \in N \bullet N \cup T$

Bemerkung: Ableitungsbäume von kontextfreien Grammatiken in Chomsky-Normalformen sind binäre Bäume

Satz F: Zu jeder kontextfreien Grammatik G existiert eine äquivalente (sprachenerzeugende) Gramatik G' in Chomsky-Normalform

Beweis:
Idee: Effektive Konstruktion in meheren Schritten:
- Entfernen von ε–Regeln
- Terminalzeichen aus rechten Seiten entfernen
- Entfernen von Kettenregeln
- Verkürzen der rechten Seite auf Länge 2

Sei $G = (N, T, P, \sigma)$ gegeben

Schritt 1: Entfernen der ε–Regeln

Bilde $M_1 := \{A \mid A\ \varepsilon \in P\} \subseteq N$

$M_{i+1} = M_i \cup \{B \mid \exists B \rightarrow u \in P \text{ mit } u \in M_i^*\} \subseteq N$ für i=1,...,n

$[X \in M_{i+1} \Rightarrow X \rightarrow^{i+1} \varepsilon]$

Es gibt: $M_1 \subseteq M_2 \subseteq ... \subseteq N$ Das heisst, die Folge der M_i ist nach oben beschränkt.
Dann gibt es ein $k \in N$ mit $M_k = M_{k+1} = ... = M_{k+r}$ für alle $r \in N$

Behauptung: $A \in M_k \Leftrightarrow A \to^* \varepsilon$ bezüglich G

"$\Rightarrow$": $A \in M_k \Rightarrow A \in M_{k-1}$ oder $\exists u \in M^*_{k-1}$ mit $A \to u \in P$
Hierdurch folgt mit vollständiger Induktion bezüglich k dass $A \to^* \varepsilon$

"$\Leftarrow$": Es gelte $A \to^* \varepsilon$. Betrachten des Ableitungsbaum.
Falls es ein „selbsteinbettendes" B gibt ($B \to^* B$), dann lasse man den schraffierten Teil weg. Es entsteht eine kürzere Ableitung $A \to^* \varepsilon$.
Eliminiere so alle selbsteinbettende Nonterminale. Der entstehende Ableitungsbaum kann höchstens Höhe $|N|$ haben.
$A \in M_{|N|} = M_k \Rightarrow$ Behauptung

Man bilde nun $Q_1 := P$

$$Q_{i+1} := \{ A \to A_1 \dots A_{j-1} A_{j+1} \dots A_m \mid A \to A_1 \dots A_{j-1} A_j A_{j+1} \dots A_m \in Q_i \text{ und } A_j \in M_k \}$$

Man entferne also schrittweise die Nonterminale aus der rechten Seite, die nach ε ableitbar sind.
Da die rechten Seiten immer kürzer werden, gibt es ein $q \in N$, so dass $Q_q = \emptyset$ und $Q_{q-1} \neq \emptyset$.

Sei $P_1' = \bigcup_{i=1}^{q-1} Q_i \setminus \{ A \to \varepsilon \mid \exists i \text{ mit } A \to \varepsilon \in Q_i \}$ (P_1' ist ohne ε-Regeln)

Beispiel: Gilt $A \to^* \varepsilon$, so nimmt man zur vorhandenen Regel $B \to uAv$ die Regel $B \to uv$
Sei $G_1' = (N, T, P, \sigma)$ und es gibt $L(G_1') = L(G) \setminus \{\varepsilon\}$.
Setze man für ein Zeichen: $\sigma \notin N$: $N_1 := N \cup \{\sigma'\}$

$$\text{Bilde } P_1 := \begin{cases} P_1' \cup \{\sigma' \to \sigma\}, \text{ falls } \varepsilon \notin L(G) \\ P_1' \cup \{\sigma' \to \sigma, \sigma \to \varepsilon\}, \text{ falls } \varepsilon \in L(G) \quad [\text{1. Anstrich in Def. E}] \end{cases}$$

Schritt 2: Terminalzeichen aus rechten Seiten entfernen
[z.B. $A \to wxw'$, $ww' \in (N \cup T)^*$, $x \in T \Rightarrow A \to wXw'$ $X \to x$]
Sei $T' := \{t' \mid t \in T\}$ eine zu T gleichmächtige Menge.
Setze $N_2 := N_1 \cup T'$ und mit $h := (N \cup T)^* \to (N \cup T')^*$ Homomorphismus

$$\text{mit } h(a) = \begin{cases} a, a \in N \\ a', a' \in T' \end{cases}$$

Bilde $P_2' = \{A \to h(a) \mid A \to u \in P_1\} \cup \{t' \to t \mid t \in T\}$
Für $G_2' = (N_2, T, P_2', \sigma')$ gilt $L(G_2') = L(G_1)$
[$A \to u \in P_2' \Rightarrow u \in T$ oder $u \in N_2^+$]

Schritt 3: Entfernen der Kettenregeln
[z.B. $A \to B$, $B \to C$, $C \to D$, wxw', $ww' \in (N \cup T)^*$, $x \in T$
$X \to wAw'$
$X \to wBw'$
$X \to wCw'$
$X \to wDw'$]

Vorgehen: Man entfernt diese Regeln, indem man sie wie in Schritt 1 in vorhergehende Regeln hineinzieht.
Für jedes $A \in N_2$ bilde
$M_1(A) = \{A\}$
$M_{i+1}(A) = M_i(A) \cup \{ B \in N_2 \mid \exists C \in M_i(A) \text{ mit } C \to B \in P_2' \} \subseteq N$
Also gilt: $\exists k$ mit $M_k(A) = M_{k+1}(A) = M_{k+r}(A)$ $(\forall r \in N)$
Und es folgt: $B \in M_k(A) \Leftrightarrow A \to^{sk-1} B$ bezüglich G_2'

[Bemerkung: Man kann ein einheitliches k für alle $A \in N_2$ wählen, da $k \subseteq |N_2|$ wegen $M_1(A) \subseteq M_2(A) \subseteq ... \subseteq N_2$]

Konstruiere $G_2 = (N_2, T, P_2, \sigma')$ mit

$P_2 = \{ A \to C_1...C_r \mid \exists A \to B_1...B_r \in P_2', r \geq 2, B_i \in N_2$ für $i = 1,...,r$
und für alle $i = 1,...,r$ gilt: $C_i \in M_k(B_i) \}$
$\cup \{ \sigma' \to u \mid \exists A \to u \in P_2'$ mit $|u| \geq 2$ und $A \in M_k(\sigma') \}$
$\cup \{ \sigma' \to t \mid t' \in M_k(\sigma'), t \in T \}$
$\cup \{ \sigma' \to \varepsilon \mid$ für $\sigma \to \varepsilon \in P_2' \}$

gilt: $L(G_2) = L(G_2') = L(G_1)$ [ohne Beweis]

Schritt 4: Verkürzung der rechten Seiten auf die Länge ≤ 2
[z.B. $A \to BCDE \Rightarrow A \to BB', B' \to CC', C' \to DE$]

Für jede Regel $A \to B_1...B_r$, $r \geq 3$, $A, B_1, ..., B_r \in N_2$ füge r-2 neue Nonterminals $A_1...A_{r-2}$ hinzu und ersetze die Regel $A \to B_1...B_r$ durch

$A \to B_1A_1$
$A_1 \to B_2A_2$
: :
: :
$A_{r-3} \to B_{r-2}A_{r-2}$
$A_{r-2} \to B_{r-1}B_r$

Beachte: Je Regel in P_2 neue Zeilen $A_1...A_{r-2}$ N" neue Nichtterminalmenge, P´neue Produktionsmenge

$P' = P_2 \setminus \{ A \to u \mid |u| \geq 3, A \to u \in P_2 \}$
$\cup \{ A \to B_1A_1, ..., A_{r-2} \to B_{r-1}B_r \mid A \to B_1, ..., B_r \in P_2, r \geq 3 \}$

Setze $N' := N'' \setminus \{ A \mid A$ tritt in P_2 nirgends auf $\}$
$G' = (N', T, P', \sigma')$ ist in Chomsky-Normalform und es gilt:
$L(G) = L(G')$

Beispiel: $G = (N, T, P, \sigma)$ mit $N = \{A, B, S\}, T = \{a, b, c\}$
$P = \{ S \to AB, A \to ab, A \to aAb, B \to c, B \to cB\}$

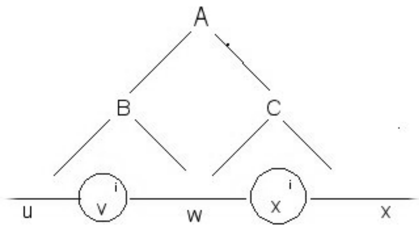
Schritt 1: keine ε-Regeln
Schritt 2: Terminalzeichen aus rechten Seiten entfernen
$T' = \{a', b', c'\}$, $N_2 = N \cup T'$
$P_2' = \{ S \to AB, A \to a'b', A \to a'Ab', B \to c', B \to c'B,$
$a' \to a, b' \to b, c' \to c\}$
Schritt 3: Kettenregeln entfernen
$M_1(B) = \{B\}$, $M_2(B) = \{B\} \cup \{ X \in N_2 \mid \exists C \in M_1(B) \text{ mit } C \to X \in P_2' \}$

<u>Schritt 4:</u> rechte Seite verkürzen $N' = N_2 \cup \{C_1\}$

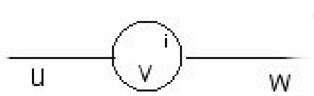
$\Rightarrow P' = \{ S \rightarrow AB, A \rightarrow a'b_1', A \rightarrow a'C_1, C_1 \rightarrow Ab', B \rightarrow c, B \rightarrow c'B, a' \rightarrow a,$
$b' \rightarrow b, c' \rightarrow c \}$

<u>Pumping-Lemma für kontextfreie Sprachen</u>

<u>Idee:</u> $A \rightarrow BC$



Korrespondenz von v und x per i



reguläre Sprache: uv^iw

<u>Satz F:</u> Zu jeder kontextfreien Sprache $C \subseteq T^*$ existiert eine Zahl $n \in N$, so dass für alle Wörter $z \in L$ mit $|z| \geq n$ gilt:
Es gibt eine Zerlegung $z = uvwxy$ mit $u,v,w,x,y \in T^*$ und

(i) $vx \neq \varepsilon$
(ii) $|vwx| \leq n$
(iii) für alle $i \in N_0$ gilt: $uv^iwx^iy \in L$