Aufgaben zum Tut am 05.12.2006

Thomas Pajor

4. Dezember 2006

Aufgabe 1.

Gegeben sei folgende kontextfreie Grammatik $G := (V, \Sigma, P, S)$ mit

$$
\begin{aligned}
P := \{S &\to aA, \\
A &\to C \mid aA \mid a, \\
B &\to C \mid bB \mid b, \\
C &\to B \mid c\}
\end{aligned}
$$

Geben Sie durch systematische Konstruktion eine äquivalente Grammatik G' in Chomsky–Normalform an.

Lösung.

Wir überführen G in Chomsky–Normalform und erhalten $G' = (V', \Sigma, P', S)$. Dazu werden wir die Produktionen und Variablen sukzessive gemäß dem Verfahren aus dem Tutorium verändern.

1. Zunächst werden Terminalzeichen aus den rechten Seiten aller Produktionen eliminiert. Dazu führen wir für jedes $x_i \in \Sigma$ ein neues Symbol X_i zu V' hinzu, und fügen in P eine Produktion $X_i \to x_i$ ein. Seien in unserem Fall X, Y, Z die neuen Variablen, die für a, b, c stehen. Dann

gilt

$$
\begin{aligned}
P_1 := \{S &\to XA,\\
A &\to C \mid XA \mid X,\\
B &\to C \mid YB \mid Y,\\
C &\to B \mid Z,\\
X &\to a,\\
Y &\to b,\\
Z &\to c\}
\end{aligned}
$$

2. Wir verkürzen alle rechten Seiten. Betrachte dazu alle Regeln der Form $A \to B_1 \cdots B_m$ mit $m > 2$. Offenbar haben wir keine solche Regeln in P_1, das heißt wir können diesen Schritt überspringen, und es gilt $P_2 := P_1$.

3. Der einzige Zyklus von Kettenregeln in P_2 ist $B \to C \to B$. Wir ersetzen also in allen Produktionen die Variable C durch B und löschen die Regel $B \to B$. Dies ergibt

$$
\begin{aligned}
P_3 := \{S &\to XA,\\
A &\to B \mid XA \mid X,\\
B &\to YB \mid Y \mid Z,\\
X &\to a,\\
Y &\to b,\\
Z &\to c\}
\end{aligned}
$$

4. Zum Schluss müssen wir noch Kettenregeln im Allgemeinen eliminieren. Da wir keine Zyklen mehr in P_3 haben, können wir die Variablen topologisch sortieren. Für zwei Variablen X_i und X_j soll gelten $X_i \to X_j \in P \Rightarrow i < j$. Dies ergibt für uns die mögliche Sortierung

$$V = \{S, A, B, Z, X, Y\}$$

Nun gehen wir für $k = n - 1 \dots 1$ die Variablen in absteigender Reihenfolge durch und suchen nach Regeln der Form $X_k \to X_l$ mit $l > k$. Gibt es so eine Regel, so wird sie gelöscht. Außerdem wird für jede Regel $X_l \to X$, wobei X beliebig ist, eine Regel $X_k \to X$ eingefügt. Dies ergibt in unserem Fall folgende Produktionen

$$
\begin{aligned}
P_4 := \{S &\to XA,\\
A &\to XA \mid a \mid b \mid c,\\
B &\to YB \mid b \mid c,\\
X &\to a,\\
Y &\to b,\\
Z &\to c\}
\end{aligned}
$$

Alle Regeln in P_4 entsprechen den Einschränkungen der Chomsky–Normalform. Mit $P' := P_4$ ist also G' eine eine zu G äquivalente Grammatik in CNF.

Aufgabe 2.

Gegeben sei folgende kontextfreie Grammatik $G := (V := \{S\}, \Sigma := \{a, b\}, P, S)$ mit

$$
\begin{aligned}
P := \{S &\to SS, && (1)\\
S &\to aSb, && (2)\\
S &\to ab\} && (3)
\end{aligned}
$$

(a) Wandeln Sie G in Chomsky–Normalform G'.

(b) Prüfen Sie mit dem Algorithmus von COCKE–YOUNGER–KASAMI ob die Wörter $w_1 := ababaabb$ und $w_2 := ababb$ in $L(G)$ enthalten sind und geben Sie ggf. eine Ableitungsfolge $S \overset{*}{\Rightarrow} w$ an.

Lösung.

(a) Überführung von G in Chomsky Normalform $G' := (V', \Sigma, P', S)$:

1. Elimination von Terminalzeichen (auf der rechten Seite). Wir führen für jedes Symbol aus Σ zwei neue Variablen Y_a und Y_b ein. Die Produktionen ändern sich wie folgt:

$$
\begin{aligned}
P_1 := \{S &\to SS \mid Y_aSY_b \mid Y_aY_b,\\
Y_a &\to b,\\
Y_a &\to b\}
\end{aligned}
$$

2. Rechte Seiten verkürzen. Die einzige Produktion in P_1, die zu lang ist, ist die Produktion $S \to Y_aSY_b$. Wir führen also eine neue Variable C ein, und ersetzen P_1 durch P_2 wie folgt:

$$
\begin{aligned}
P_2 := \{S &\to SS \mid Y_aC \mid Y_aY_b,\\
C &\to SY_b,\\
Y_a &\to a,\\
Y_b &\to b\}
\end{aligned}
$$

3. Da wir keine Kettenproduktionen haben, sind wir fertig, und es gilt $G' = (\{S, Y_a, Y_b, C\}, \Sigma, P_2, S)$ ist in Chomsky–Normalform.

(b) Anwendung des Algorithmus von COCKE–YOUNGER–KASAMI ergibt folgende Tabellen:

Prüfung ob $w_1 \in L(G)$:

	a	b	a	b	a	a	b	b
1	Y_a	Y_b	Y_a	Y_b	Y_a	Y_a	Y_b	Y_b
2	S	-	S	-	-	S	-	
3	-	-	-	-	-	C		
4	S	-	-	-	S			
5	-	-	-	-				
6	-	-	S					
7	-	-						
8	S							

Der Übersicht wegen, habe ich hier nicht die Mengenschreibweise verwendet. Korrekterweise muss jeder Eintrag in der Tabelle eine Menge sein, da es sein kann dass ein Eintrag mehrere Elemente enthalten kann (ist hier nicht der Fall). Formal korrekt wäre auch $\emptyset$ statt „-“ zu schreiben.

Da der Eintrag $(8, 1)$ das Symbol S enthält, lässt sich w_1 aus S ableiten, und somit gilt $w_1 \in L(G)$.

Eine gültige Ableitungsfolge ist:

$$S \Rightarrow SS \Rightarrow SSS \Rightarrow Y_aY_bSS \Rightarrow Y_aY_bYaY_bS \Rightarrow Y_aY_bY_aY_bY_aC$$
$$\Rightarrow Y_aY_bY_aY_bY_aSY_b \Rightarrow Y_aY_bY_aY_bY_aY_aY_bY_b \overset{*}{\Rightarrow} ababaabb$$

Prüfung ob $w_2 \in L(G)$:

	a	b	a	b	b
1	Y_a	Y_b	Y_a	Y_b	Y_b
2	S	-	S	-	
3	-	-	-		
4	S	-			
5	C				

Der Eintrag $(5, 1)$ enthält nicht das Startsymbol S. Somit lässt sich w_2 nicht aus S ableiten, und es gilt $w_2 \notin L(G)$.

Aufgabe 3.

Zeigen oder widerlegen Sie: Die Sprache

$$L := \{ww \mid w \in \{0, 1\}^*\}$$

ist kontextfrei.

Lösung.

Wir vermuten, dass die Aussage nicht gilt, und möchten dies mit dem Pumping Lemma für kontextfreie Sprachen zeigen.

Sei also n die Zahl aus dem Pumping Lemma. Betrachte zu jedem n das Wort $z := 0^n 1^n 0^n 1^n \in L$. Offenbar ist $|z| > n$ stets erfüllt. Betrachte nun jede Zerlegung $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$. Wegen $|vx| \geq 1$ und $|vwx| \leq n$ folgt $|w| \leq n - 1$. Das heißt es gilt immer $u \neq x$, und damit ist

$$uv^0 wx^0 y = uwy \neq L$$

Mit dem Pumping Lemma folgt dass L nicht kontextfrei ist. $\square$