Einführung in die Computerlinguistik
Hausaufgabe 7, Abgabe 11.06.2012

Laura Kallmeyer

SS 2012, Heinrich-Heine-Universität Düsseldorf

Aufgabe 1 *Gegeben ist die CFG G_1 mit den Nichtterminalen $\{S, T, A, B\}$, den Terminalen $\{a, b\}$, dem Startsymbol S und den Produktionen*

$$\begin{array}{lll} S \to ATA & S \to BTB & \\ T \to ATA & T \to BTB & T \to \varepsilon \\ A \to a & B \to b & \end{array}$$

1. *Transformieren Sie G_1 in eine äquivalente CFG G_1', die keine ε-Produktionen enthält.*
2. *Transformieren Sie G_1' in eine äquivalente CFG G_1'' in Chomsky Normal Form.*

Lösung:

1. Berechnen Sie das Set N_ϵ: $N_\epsilon = \{T\}$
   Daraus ergeben sich folgende Produktionen für $G_1^{'}$:
   $S \to ATA \quad S \to BTB \quad S \to AA \quad S \to BB$
   $T \to ATA \quad T \to BTB \quad T \to AA \quad T \to BB$
   $A \to a \quad B \to b$
2. Für die Transformation in die CNF erstellen wir die neuen Nichtterminalen C_1, C_2. Daraus ergeben sich für $G_1^{''}$ folgende Produktionen:
   $S \to AC_1 \quad S \to BC_2 \quad S \to AA \quad S \to BB$
   $T \to AC_1 \quad T \to BC_2 \quad T \to AA \quad T \to BB$
   $C_1 \to TA \quad C_2 \to TB \quad A \to a \quad B \to b$

Aufgabe 2

Fassen Sie die Grundidee des Beweises des Pumping Lemmas für kontextfreie Sprachen in maximal 150 Wörter zusammen; verzichten Sie bitte weitestgehend auf Formeln und verwenden Sie maximal eine Skizze.

Aufgabe 3 *Betrachten Sie folgende Sprachen:*

- $L_1 = \{a^n b^n a^n b^n \mid n > 0\}$
- $L_2 = \{a^n b^n c^n d^n \mid n > 0\}$
- $L_3 = \{ww | w \in \{a, b\}^+\}$

1. *Zeigen Sie mit Hilfe des Pumping Lemmas , dass L_1 keine kontextfreie Sprache ist. Sie können sich dabei an dem Beweis für $\{a^n b^n c^n \mid n \geq 1\}$ orientieren.*
2. *Zeigen Sie anschließend, dass L_2 ebenfalls nicht kontextfrei ist. Hierbei können Sie sich auf das unter 1. gezeigte Resultat und die Abgeschlossenheit kontextfreier Sprachen unter Homomorphismen beziehen.*

3. *Zeigen Sie nun, dass auch L_3 nicht kontextfrei ist. Hierbei können Sie wieder das Resultat aus 1. verwenden in Verbindung mit der Abgeschlossenheit kontextfreier Sprachen unter Schnittbildung mit regulären Sprachen.*

Lösung:

1. Zu zeigen: L_1 ist nicht kontextfrei.

   Annahme: L_1 ist kontextfrei. Dann erfüllt L_1 das Pumping Lemma mit einer Konstanten k. Insbesondere muss es also in $w = a^k b^k a^k b^k$ Teilstrings v_1, v_2 geben, die iteriert werden können, und von denen mindestens einer nicht leer ist. v_1 und v_2 sind entweder so, dass einer von ihnen nur as enthält und der andere gleichviel bs. Dann würde aber die Iteration zu Wörtern führen, bei denen die vier Teile a^n, b^n, a^n, b^n nicht mehr alle vier gleich lang sind. Oder mindestens einer der Strings v_1 und v_2 enthält sowohl as als auch bs. Dein entsteht aber bei Iteration notwendig ein unzulässiges Mischen von as und bs.

   $\Rightarrow$ L_1 erfüllt nicht das Pumping Lemma, unsere Annahme, dass L_1 kontextfrei ist, führt also zu einem Widerspruch und ist somit falsch.

2. Zu zeigen: L_2 ist nicht kontextfrei.

   Annahme: L_2 ist kontextfrei. Dann muss auch das Bild von L_2 unter dem Homomorphismus f mit $f(a) = f(c) = a, f(b) = f(d) = b$ kontextfrei sein. Dieses Bild ist die Sprache L_1, von der wir gerade gezeigt haben, dass sie nicht kontextfrei ist. Unsere Annahme, dass L_2 kontextfrei ist, führt also zu einem Widerspruch und ist somit falsch.

3. Zu zeigen: L_3 ist nicht kontextfrei.

   Annahme: L_3 ist kontextfrei. Dann muss auch der Schnitt von L_3 mit der regulären Sprache $a^*b^*a^*b^*$ kontextfrei sein. Dieser Schnitt ist die Sprache L_1, von der wir wissen, dass sie nicht kontextfrei ist. Unsere Annahme, dass L_3 kontextfrei ist, führt also zu einem Widerspruch und ist somit falsch.