Formale Grundlagen der Informatik I
6. Übungsblatt

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik
Prof. Dr. Ulrich Kohlenbach
Alexander Kreuzer
Pavol Safarik

SS 2012

Gruppenübung

Aufgabe G1
Betrachten Sie die kontextfreie Sprache L, die von der folgenden Grammatik G erzeugt wird:

$$G: \quad X_0 \quad \to \quad ab \mid ba \mid X_0X_0 \mid aX_0b \mid bX_0a$$

(a) Beschreiben Sie L umgangssprachlich.

(b) Bringen Sie G in Chomsky-Normalform.

(c) Wenden Sie den CYK Algoritmus an, um zu bestimmen, ob $bbab \in L$ und $aabbab \in L$.

(d) Konstruieren Sie einen Kellerautomaten, welcher L erkennt.

Lösungsskizze:

(a) L besteht aus den Wörtern in $\{a, b\}^*$, die nicht leer sind und genauso viele a wie b enthalten.

Klar ist, dass alle Wörter w, die man ableiten kann, genauso viele a wie b enthalten und dass das leere Wort nicht ableitbar ist. Andererseits kann man das Argument in Aufgabe G1 auf dem 5. Übungsblatt einfach zu einem Beweis umschreiben, dass alle nicht-leeren Wörter mit genauso vielen a wie b in G ableitbar sind.

(b) Es gibt keine Regeln der Form $X \to Y$ oder $X \to \varepsilon$, deshalb können wir direkt damit beginnen, Buchstaben durch Variablen zu ersetzen und erhalten:

$$\begin{array}{lcl} X_0 & \to & Z_aZ_b \mid Z_bZ_a \mid X_0X_0 \mid Z_aX_0Z_b \mid Z_bX_0Z_a \\ Z_a & \to & a \\ Z_b & \to & b \end{array}$$

Nun werden die Regeln der Form $X \to Y_1Y_2 \dots Y_n$ mit $n \geq 3$ aufgelöst und wir erhalten als Endergebnis:

$$\begin{array}{lcl} X_0 & \to & Z_aZ_b \mid Z_bZ_a \mid X_0X_0 \mid Z_aX \mid Z_bY \\ X & \to & X_0Z_b \\ Y & \to & X_0Z_a \\ Z_a & \to & a \\ Z_b & \to & b \end{array}$$

(c) Erzeugbare Teilworte der Länge 1:

$$\begin{array}{lcl} a & : & Z_a \\ b & : & Z_b \end{array}$$

Erzeugbare Teilworte der Länge 2:

$$\begin{array}{rcl} aa & : & - \\ ab & : & X_0 \\ bb & : & - \\ ba & : & X_0 \end{array}$$

Erzeugbare Teilworte der Länge 3:

$$\begin{array}{rcl} aab & : & - \\ abb & : & X \\ bba & : & - \\ bab & : & X \end{array}$$

Erzeugbare Teilworte der Länge 4:

$$\begin{array}{rcl} aabb & : & X_0 \\ abba & : & X_0 \\ bbab & : & - \end{array}$$

Erzeugbare Teilworte der Länge 5:

$$\begin{array}{rcl} aabba & : & Y \\ abbab & : & X \end{array}$$

Deshalb ist $bbab$ nicht erzeugbar und $aabbab$ ist erzeugbar (z.B. mittels $X_0 \to Z_a X$).

(d) Sei $\mathcal{P} = (\Sigma, Q, q_0, \Delta, A, \Gamma, \#)$ der Kellerautomat mit Eingabealphabet $\Sigma = \{a, b\}$, Zustandsmenge $Q = \{q_0, q_1\}$, q_0 als Anfangszustand, $A = \{q_1\}$ als Menge der akzeptierenden Zustände, Kelleralphabet $\Gamma = \{\#, A, B\}$ und Übergangsrelation Δ gegeben durch

$$\begin{array}{lll} \{ & (q_0, \#, a, A\#, q_1) & \\ & (q_0, \#, b, B\#, q_1) & \\ & (q_1, A, a, AA, q_1) & \\ & (q_1, A, b, \varepsilon, q_1) & \\ & (q_1, B, a, \varepsilon, q_1) & \\ & (q_1, B, b, BB, q_1) & \\ & (q_1, \#, a, A\#, q_1) & \\ & (q_1, \#, b, B\#, q_1) & \\ & (q_1, \#, \varepsilon, \varepsilon, q_1) & \}. \end{array}$$

$\mathcal{P}$ erkennt die Sprache L. Der Automat ist in Zustand q_1 genau dann, wenn ein nicht-leeres Wort bereits gelesen wurde. Außerdem wird der Keller benutzt um die Differenz der a und b im bisher gelesenen Wort zu zählen: Wenn bisher das Wort w gelesen wurde und $|w|_a > |w|_b$ dann befinden sich genau $|w|_a - |w|_b$ viele A im Speicher, falls $|w|_b > |w|_a$ gilt die analoge Aussage für B.

Alternativ kann man auch die Konstruktion aus dem Beweis des Satzes 4.1.5 aus dem Skript anwenden, die eine beliebige kontextfreie Grammatik in einen Kellerautomaten übersetzt, der die von der Grammatik erzeugte Sprache akzeptiert.

Aufgabe G2

(a) Zeigen Sie, dass die Sprache

$$L = \{a^n b^m a^n b^m \,:\, m, n \geq 0\}$$

nicht kontextfrei ist.

(b) Sei L eine kontextfreie Sprache und M eine reguläre Σ-Sprache. Zeigen Sie, dass $L \cap M$ eine kontextfreie Σ-Sprache ist.

Hinweis: Sei $\mathcal{P}$ ein Kellerautomat für L, und $\mathcal{A}$ ein NFA für M. Konstruieren Sie daraus (wie in Lemma 2.2.11(a) im Skript) einen Kellerautomat $\mathcal{Q}$, der $L \cap M$ erkennt.

(c) Benutzen Sie Aufgabenteilen (a) und (b) um zu schließen, dass

$$N = \{ww \,:\, w \in \{a,b\}^*\}$$

keine kontextfreie Sprache ist.

(d) Zusatzaufgabe: Warum kann man nicht 2 Kellerautomaten parallel simulieren? Warum ist der Durchschnitt kontextfreier Sprachen im allgemeinen nicht kontextfrei?

Lösungsskizze:

(a) Wir zeigen, dass die Bedingung des Pumping Lemmas verletzt ist. Das heißt:

Für jedes $i \in \mathbb{N}$ gibt es ein $x \in L$ mit $|x| \geq i$, so dass für alle Zeichenreihen y, u, v, w, z, mit $x = y \cdot u \cdot v \cdot w \cdot z$, $|uw| > 0$ und $|uvw| \leq i$, es ein $j \in \mathbb{N}$ gibt mit

$$y \cdot u^j \cdot v \cdot w^j \cdot z \notin L.$$

Um dies zu zeigen, sei $i \in \mathbb{N}$ beliebig und

$$x = a^i b^i a^i b^i.$$

Offensichtlich ist $x \in L$ und $|x| \geq i$. Seien also y, u, v, w, z, mit $x = y \cdot u \cdot v \cdot w \cdot z$, $|uw| > 0$ und $|uvw| \leq i$. Wir wählen $j = 2$ und behaupten

$$x' = y \cdot u^2 \cdot v \cdot w^2 \cdot z \notin L.$$

Beweis: Wir bezeichnen die Teilwörter der Form a^i und b^i in x als *Blöcke*. Es gibt jetzt drei Möglichkeiten:

- Entweder u oder w überschreitet eine Blockgrenze (beides zusammen ist wegen $|uvw| \leq i$ unmöglich!). Dann enthält entweder u oder w sowohl a als b. Das bedeutet, dass x' nicht die Form $a^*b^*a^*b^*$ hat und deshalb nicht in L enthalten sein kann.
- u und w sind in demselben Block enthalten. Dann hat x' einen Block der länger ist als die andere drei und ist deshalb nicht in L enthalten.
- u und w sind in verschiedenen Blöcken. Diese Blöcke müssen benachbart sein, da $|uvw| \leq i$. Das heißt, dass entweder u nur aus a besteht und w aus b, oder umgekehrt. Dann haben die zwei a-Blöcke in x' verschiedene Länge (wenn uw a enthält), oder die zwei b-Blöcke (wenn uw b enthält; beides ist natürlich auch möglich!). Jedenfalls ist x' nicht in L enthalten.

(b) Sei $\mathcal{P} = (\Sigma, Q^1, q_0^1, \Delta^1, A^1, \Gamma, \#)$ ein Kellerautomat, der L erkennt, und $\mathcal{A} = (\Sigma, Q^2, q_0^2, \Delta^2, A^2)$ ein NFA der M erkennt. Wir konstuieren einen Kellerautomat $\mathcal{Q}$, der $\mathcal{P}$ und $\mathcal{A}$ parallel simuliert analog zur Bildung von Produktautomaten für den Schnitt regulärer Sprachen. Sei $\mathcal{Q} := (\Sigma, Q, q_0, \Delta, A, \Gamma, \#)$, wobei:

$$\begin{aligned} Q &= Q^1 \times Q^2 \\ q_0 &= (q_0^1, q_0^2) \\ A &= A^1 \times A^2, \end{aligned}$$

und

$$\begin{aligned} ((q,p), \gamma, x, \beta, (q',p')) \in \Delta \quad :\Leftrightarrow \quad & \big((q, \gamma, x, \beta, q') \in \Delta^1 \text{ und } (p, x, p') \in \Delta^2 \text{ und } x \neq \varepsilon\big) \text{ oder} \\ & \big((q, \gamma, x, \beta, q') \in \Delta^1 \text{ und } p = p' \text{ und } x = \varepsilon\big) \end{aligned}$$

(c) Sei L die Sprache aus Aufgabe (a) und $M := L(a^*b^*a^*b^*)$. M ist regulär und es gilt $L = M \cap N$. Wäre N also kontextfrei, dann wäre nach Aufgabe (b) L auch kontextfrei. Wir haben aber in Aufgabe (a) gezeigt, dass L nicht kontextfrei ist.

(d) Folgt aus (c).

Hausübung

Aufgabe H1 (6 Punkte)

Betrachten Sie nun die folgende Grammatik $G_b = (\Sigma, V_b, P_b, S)$ mit $\Sigma = \{a, b\}$ und $V_b = \{S, T, X, Y\}$ und den folgenden Produktionen P_b:

$$\begin{aligned} S &\to TS \mid XS \mid a \\ T &\to TT \mid YY \mid YS \\ X &\to a \\ Y &\to b \end{aligned}$$

Entscheiden Sie mit Hilfe des CYK-Algorithmus, ob das Wort $abbbaa$ in der von G_b erzeugten Sprache $L(G_b)$ enthalten ist.