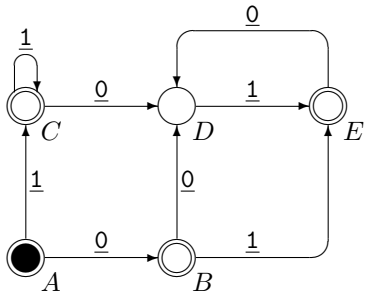
Theoretische Informatik und Logik
Übungsblatt 3 (WS 2011)
Lösungen

Aufgabe 3.1 Sei $G = \langle \{S, T\}, \{\underline{0}, \underline{1}\}, \{S \to S\underline{0}\underline{1} \mid \underline{0} \mid T, \quad T \to \underline{1}T \mid \varepsilon\}, S\rangle$
Geben Sie eine reguläre Grammatik an, welche dieselbe Sprache wie G erzeugt.

Lösung Wir überlegen zunächst, welche Sprache von G erzeugt wird. Nach kurzem Nachdenken stellen wir fest, dass

$$\mathcal{L}(G) = (\{\underline{1}\}^* \cup \{\underline{0}\})\{\underline{0}\underline{1}\}^*$$

ist. $\mathcal{L}(G)$ ist also eine reguläre Sprache, die z.B. von folgendem DEA akzeptiert wird:



Daraus können wir direkt folgende reguläre Grammatik ableiten, welche die von G erzeugte Sprache erzeugt:

$$G' = (\{A, B, C, D, E\}, \{\underline{0}, \underline{1}\}, P, A)$$

wobei

$$P = \{A \to \underline{0}B \mid \underline{1}C \mid \varepsilon, \quad B \to \underline{0}D \mid \underline{1}E \mid \varepsilon, \quad C \to \underline{1}C \mid \underline{0}D \mid \varepsilon, \quad D \to \underline{1}E, \quad E \to \underline{0}D \mid \varepsilon\}$$

Aufgabe 3.2 Sei $z(x)$ die durch die Darstellung x repräsentierte Binärzahl und

$$L_{ADD} = \{u \underline{+} v \underline{=} w \mid u, v, w \in \{\underline{0}, \underline{1}\}^* \text{ und } z(u) + z(v) = z(w)\}$$

(L_{ADD} enthält also z.B. die Wörter $\underline{1}\underline{1}\underline{+}\underline{1}\underline{=}\underline{1}\underline{0}\underline{0}$ und $\underline{1}\underline{0}\underline{0}\underline{+}\underline{1}\underline{1}\underline{=}\underline{1}\underline{1}\underline{1}$, hingegen ist z.B. $\underline{1}\underline{1}\underline{+}\underline{1}\underline{=}\underline{0}\underline{1}$ nicht in L_{ADD}.)

Zeigen Sie mit Hilfe des Pumping Lemmas für reguläre Sprachen, dass L_{ADD} nicht regulär ist.

Lösung Beweis indirekt. Angenommen, L ist regulär. Sei dann m die Konstante aus dem Pumping Lemma und das gewählte Wort z.B.

$$w = \underline{1}^m \underline{+} \underline{0} \underline{=} \underline{1}^m$$

Dann gilt $w \in L$ und $|w| = 2m + 3 > m$.
Nach dem Pumping Lemma kann w in xyz so aufgeteilt werden, dass $|xy| \leq m$, $|y| > 0$ und für alle $i \geq 0$ gilt: $xy^iz \in L$.
Nachdem $|xy| \leq m$ und $|y| > 0$, kann kann xy nur aus Symbolen $\underline{1}$ bestehen, wobei $y = \underline{1}^k$ ist für $1 \leq k \leq m$.
Wählen wir nun z.B. $i = 0$, so müsste auch $\underline{1}^{m-k} \underline{+} \underline{0} \underline{=} \underline{1}^m$ aus L sein, was aber nicht der Fall ist! (Denn offensichtlich ist $z(\underline{1}^m)$ nicht die Summe von $z(\underline{1}^{m-k}) + z(\underline{0})$.) Wir haben also einen Widerspruch gefunden, L kann somit keine reguläre Sprache sein.

Aufgabe 3.3 Geben Sie jeweils eine kontextfreie Grammatik an, die folgende Sprachen erzeugt. Geben Sie jeweils auch einen Ableitungsbaum für ein Wort $w \in L(G_i)$, $1 \leq i \leq 3$, mit $|w| \geq 6$ an.

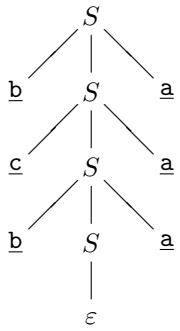
a) $\mathcal{L}(G_1) = \{w\underline{\mathtt{a}}^{|w|} \mid w \in \{\underline{\mathtt{b}}, \underline{\mathtt{c}}\}^*\}$

b) $\mathcal{L}(G_2) = \{w \in \{\underline{\mathtt{a}}, \underline{\mathtt{b}}\}^* \mid |w|_{\underline{\mathtt{a}}} = |w|_{\underline{\mathtt{b}}}\}$

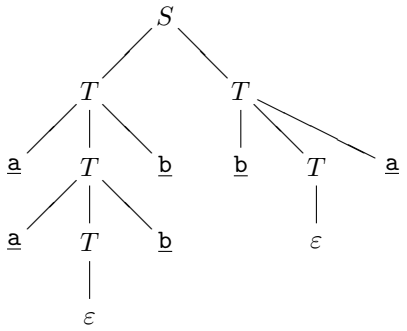
c) $\mathcal{L}(G_3) = \{\underline{\mathtt{c}}^k\underline{\mathtt{d}}^l \mid k \geq 1, l \geq 2k-1\}$

Lösung

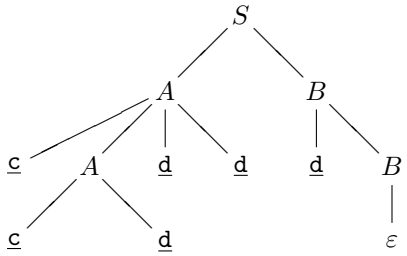
a) $G_1 = \langle\{S\}, \{\underline{\mathtt{a}}, \underline{\mathtt{b}}, \underline{\mathtt{c}}\}, \{S \to \underline{\mathtt{b}}S\underline{\mathtt{a}} \mid \underline{\mathtt{c}}S\underline{\mathtt{a}} \mid \varepsilon\}, S\}\rangle$
Ableitungsbaum für $w = \underline{\mathtt{b}}\,\underline{\mathtt{c}}\,\underline{\mathtt{b}}\,\underline{\mathtt{a}}\,\underline{\mathtt{a}}\,\underline{\mathtt{a}}$:



b) $G_2 = \langle\{T\}, \{\underline{\mathtt{a}}, \underline{\mathtt{b}}\}, \{T \to \underline{\mathtt{a}}T\underline{\mathtt{b}} \mid \underline{\mathtt{b}}T\underline{\mathtt{a}} \mid TT \mid \varepsilon\}, T\}\rangle$
Ableitungsbaum für $w = \underline{\mathtt{a}}\,\underline{\mathtt{a}}\,\underline{\mathtt{b}}\,\underline{\mathtt{b}}\,\underline{\mathtt{b}}\,\underline{\mathtt{a}}$:



c) $G_3 = \langle\{S, A, B\}, \{\underline{\mathtt{c}}, \underline{\mathtt{d}}\}, \{S \to AB, A \to \underline{\mathtt{c}}A\underline{\mathtt{d}}\,\underline{\mathtt{d}} \mid \underline{\mathtt{c}}\,\underline{\mathtt{d}}, B \to \underline{\mathtt{d}}B \mid \varepsilon\}, S\}\rangle$
Ableitungsbaum für $w = \underline{\mathtt{c}}\,\underline{\mathtt{c}}\,\underline{\mathtt{d}}\,\underline{\mathtt{d}}\,\underline{\mathtt{d}}\,\underline{\mathtt{d}}$:



Aufgabe 3.4 Gegeben ist die folgende Grammatik G:

$$\begin{aligned} G &= \langle\{S, A, B\}, \{\underline{0}, \underline{1}\}, P, S\rangle \text{ mit} \\ P &= \{S \to ASB \mid \varepsilon, \quad A \to \underline{0}AS \mid \underline{0}, \quad B \to S\underline{1}S \mid A \mid \underline{1}\,\underline{1}\}. \end{aligned}$$

Transformieren Sie G in Chomsky Normalform.

Lösung

- **Schritt 1 und 2** *(Elimination nutzloser Produktionen)*: Alle Symbole sind von S aus erreichbar, aus jeder Variablen ist ein Terminalwort ableitbar. Es gibt in G also keine nutzlosen Produktionen. Wir erhalten entsprechend $G_1 = G_2 = G$ und es gilt offensichtlich $\mathcal{L}(G_1) = \mathcal{L}(G_2) = \mathcal{L}(G)$.

- **Schritt 3** *(Elimination der ε-Produktionen)*: Wir entfernen die Produktion $S \to \varepsilon$ (es gilt also $\varepsilon \in \mathcal{L}(G)$!) und erhalten dementsprechend einige neue Produktionen, da S nun überall dort, wo es auf der rechten Seite einer Produktion vorkommt, eliminiert wird (d.h., gleich ε gesetzt wird) oder nicht. Nachdem dadurch keine nutzlosen Produktionen eingeführt wurden, erhalten wir

$$\begin{aligned} G_3 &= \langle \{S, A, B\}, \{\underline{0}, \underline{1}\}, P_3, S\rangle \text{ mit} \\ P_3 &= \{S \to ASB \mid AB, \quad A \to \underline{0}AS \mid \underline{0}A \mid \underline{0}, \quad B \to S\underline{1}S \mid S\underline{1} \mid \underline{1}S \mid \underline{1} \mid A \mid \underline{1}\underline{1}\} \end{aligned}$$

  Und es gilt: $\mathcal{L}(G_3) = \mathcal{L}(G) - \{\varepsilon\}$.

- **Schritt 4** *(Elimination der Einheits-Produktionen)*: Wir entfernen nun die Einheitsproduktion $B \to A$ (wodurch keine nutzlosen Produktionen eingeführt wurden) und erhalten G_4 mit $\mathcal{L}(G_4) = \mathcal{L}(G) - \{\varepsilon\}$:

$$\begin{aligned} G_4 &= \langle \{S, A, B\}, \{\underline{0}, \underline{1}\}, P_4, S\rangle \text{ mit} \\ P_4 &= \{S \to ASB \mid AB, \quad A \to \underline{0}AS \mid \underline{0}A \mid \underline{0}, \quad B \to S\underline{1}S \mid S\underline{1} \mid \underline{1}S \mid \underline{1} \mid \underline{1}\underline{1} \mid \underline{0}AS \mid \underline{0}A \mid \underline{0}\} \end{aligned}$$

- **CNF**: Um von der reduzierten Grammatik G_4 auf eine äquivalente Grammatik G' in CNF zu kommen, führen wir noch folgende Schritte durch:
  - Wir ersetzen die Symbole $\underline{0}$ bzw. $\underline{1}$ (soferne diese nicht alleine auf der rechten Seite einer Produktion vorkommen) durch entsprechende Nonterminale $X_{\underline{0}}$ bzw. $X_{\underline{1}}$ und fügen die Produktionen $X_{\underline{0}} \to \underline{0}$ und $X_{\underline{1}} \to \underline{1}$ hinzu:

$$\begin{aligned} P'' = \{ \quad & S \to ASB \mid AB, \\ & A \to X_{\underline{0}}AS \mid X_{\underline{0}}A \mid \underline{0}, \\ & B \to SX_{\underline{1}}S \mid SX_{\underline{1}} \mid X_{\underline{1}}S \mid X_{\underline{1}}X_{\underline{1}} \mid \underline{1} \mid X_{\underline{0}}AS \mid X_{\underline{0}}A \mid \underline{0}, \\ & X_{\underline{1}} \to \underline{1}, \quad X_{\underline{0}} \to \underline{0} \quad \} \end{aligned}$$

  - Nun gibt es noch einige Produktionen, die auf der rechten Seite mehr als zwei Nonterminale haben. Diese ersetzen wir folgendermaßen:

$$\begin{array}{lll} S \to ASB & \text{durch} & S \to AY_1, \quad Y_1 \to SB \\ A \to X_{\underline{0}}AS & \text{durch} & A \to X_{\underline{0}}Y_2, \quad Y_2 \to AS \\ B \to SX_{\underline{1}}S & \text{durch} & B \to SY_3, \quad Y_3 \to X_{\underline{1}}S \\ B \to X_{\underline{0}}AS & \text{durch} & B \to X_{\underline{0}}Y_2 \end{array}$$

  - Nachdem $\varepsilon \in \mathcal{L}(G)$ und S auf der rechten Seite von Produktionen vorkommt, fügen wir noch ein neues Startsymbol S' zusammen mit den Produktionen $S' \to AY_1 \mid AB \mid \varepsilon$ hinzu, und erhalten schließlich die zu G äquivalente Grammatik G' in CNF mit $\mathcal{L}(G') = \mathcal{L}(G)$:

$$\begin{aligned} G' = \langle \quad & \{S', S, A, B, X_{\underline{0}}, X_{\underline{1}}, Y_1, Y_2, Y_3\}, \{\underline{0}, \underline{1}\}, P', S' \quad \rangle \text{ mit} \\ P' = \{ \quad & S' \to AY_1 \mid AB \mid \varepsilon, \\ & S \to AY_1 \mid AB, \\ & A \to X_{\underline{0}}Y_2 \mid X_{\underline{0}}A \mid \underline{0}, \\ & B \to SY_3 \mid SX_{\underline{1}} \mid X_{\underline{1}}S \mid X_{\underline{1}}X_{\underline{1}} \mid \underline{1} \mid X_{\underline{0}}Y_2 \mid X_{\underline{0}}A \mid \underline{0}, \\ & X_{\underline{1}} \to \underline{1}, \\ & X_{\underline{0}} \to \underline{0}, \\ & Y_1 \to SB, \\ & Y_2 \to AS, \\ & Y_3 \to X_{\underline{1}}S \quad \} \end{aligned}$$