Formale Grundlagen der Informatik I
6. Übungsblatt

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik
Prof. Dr. Martin Ziegler
Davorin Lešnik, Ph.D.
Carsten Rösnick

Sommersemester 2013
20. 05. 2013

Gruppenübung

Aufgabe G13 (Kellerautomaten)
Welche Sprache wird durch den nachfolgenden Kellerautomaten (KA) beschrieben?
Es sei $\mathscr{P} = (\Sigma, Q, q, \Delta, A, \Gamma, \#)$ mit $Q = \{q\}$, $\Gamma = \{\#, |\}$, $A = \{q\}$ und Transitionen

$$
\begin{array}{lllllll}
\{ & (q, & \#, & \varepsilon, & \varepsilon, & q) & \\
 & (q, & \#, & b, & \#, & q) & \\
 & (q, & \#, & a, & |\#, & q) & \\
 & (q, & |, & a, & ||, & q) & \\
 & (q, & |, & b, & \varepsilon, & q) & \\
 & (q, & \#, & c, & \#, & q) & \\
 & (q, & |, & c, & |, & q) & \}.
\end{array}
$$

Lösung: Die Sprache $\mathrm{L}(\mathscr{P})$ enthält alle $w \in \Sigma^*$, sodass zu jedem a man eine spätere Stelle mit einem b finden kann derart, dass jedes b zu höchstens einem a gehört.

Aufgabe G14 (Kellerautomaten und kontextfreie Grammatiken)
Sei $L := \left\{a^i b^j c^k \,\middle|\, i = j + k\right\}$.

(a) Konstruieren Sie einen Kellerautomaten für L.

(b) Geben Sie eine kontextfreie Grammatik für L an.

Lösung:

(a) Sei $\mathscr{P} = (\Sigma, Q, q_a, \Delta, A, \Gamma, \#)$ der Kellerautomat mit Eingabealphabet $\Sigma = \{a, b, c\}$, Zustandsmenge $Q = \{q_a, q_b, q_c\}$, q_a als Anfangszustand, $A = \{q_a, q_b, q_c\}$ als Menge der akzeptierenden Zustände, Kelleralphabet $\Gamma = \{\#, |\}$ und Übergangsrelation Δ gegeben durch

$$
\begin{array}{lllllll}
\{ & (q_a, & \#, & \varepsilon, & \varepsilon, & q_a) & \\
 & (q_a, & \#, & a, & |, & q_a) & \\
 & (q_a, & |, & a, & ||, & q_a) & \\
 & (q_a, & |, & b, & \varepsilon, & q_b) & \\
 & (q_a, & |, & c, & \varepsilon, & q_c) & \\
 & (q_b, & |, & b, & \varepsilon, & q_b) & \\
 & (q_b, & |, & c, & \varepsilon, & q_c) & \\
 & (q_c, & |, & c, & \varepsilon, & q_c) & \}.
\end{array}
$$

Dann erkennt $\mathscr{P}$ die Sprache L.

(b) Wir gehen die Konstruktion einer $\mathrm{L}(\mathscr{P})$ erkennenden Grammatik G (wie im Beweis von Satz 4.1.5 im Skript) systematisch an und teilen diese insbesondere die im Satz angegebenen vier Schritte auf. Dabei seien stets $q', q'' \in Q$:

1.
$$X_0 \to (q_a, \#, q')$$

2.
$$\begin{aligned}(q_a, \#, q_a) &\to \varepsilon\\ (q_a, |, q_b) &\to b\\ (q_a, |, q_c) &\to c\\ (q_b, |, q_b) &\to b\\ (q_b, |, q_c) &\to c\\ (q_c, |, q_c) &\to c\end{aligned}$$

3.
$$(q_a, \#, q') \to a(q_a, |, q')$$

4.
$$(q_a, |, q'') \to a(q_a, |, q')(q', |, q'')$$

Aufgabe G15 (Pumping-Lemma und Abschlusseigenschaften der Typ-2 Sprachklasse)
Sei $\Sigma := \{a, b\}$. Beweisen oder widerlegen Sie die folgenden Aussagen.

(a) Die Sprache $L_1 := \{t\,t \mid t \in \Sigma^*\}$ ist kontextfrei.
Bemerkung: Zum Vergleich: Die Sprache der Palindrome, $L_0 := \{t\,t^{-1} \mid t \in \Sigma^*\}$, ist kontextfrei.

(b) Die Sprache $L_2 := \{s\,t\,s^{-1} \mid s, t \in \Sigma^*,\ |s| = |t|\}$ ist kontextfrei.

(c) Der Schnitt einer kontextfreien mit einer regulären Sprache ist wieder kontextfrei.

(d) Der Schnitt einer kontext*sensitiven* mit einer regulären Sprache ist kontext*frei*.

Lösung:

(a) L_1 ist nicht kontextfrei. Beweis durch Pumping-Lemma mit Wörtern $x_n := a^n b^n a^n b^n$.

(b) Auch L_2 ist nicht kontextfrei. Der Beweis kann wieder mittels Pumping-Lemmas mit Wörtern $x_n := a^n b^n a^n b^n b^n a^n$ geführt werden.

(c) Sei L kontextfrei und L' regulär. Die Idee: Konstruiere einen Produktautomaten, der *parallel* die beiden ursprünglichen Automaten simuliert. Formal: Sei $\mathscr{P} = (\Sigma, Q, q_0, \Delta, A, \Gamma, \#)$ ein KA für L und $\mathscr{A}' = (\Sigma, Q', q_0', \delta', A')$ ein DEA für L'. Ein KA $\mathscr{P}'' = (\Sigma, Q'', q_0'', \Delta'', A'', \Gamma'', \#)$ für $L'' := L \cap L'$ ist dann wie folgt zu konstruieren:

$$\begin{aligned}Q'' &:= Q \times Q'\\ q_0'' &:= (q_0, q_0')\\ \Delta'' &:= \Big\{\Big((q, q'), \gamma, x, \beta, \big(p, \delta'(q', x)\big)\Big) \;\Big|\; (q, \gamma, x, \beta, p) \in \Delta,\ q' \in Q',\ x \in \Sigma\Big\}\\ &\quad \cup \;\Big\{\Big((q, q'), \gamma, \varepsilon, \beta, \big(p, q'\big)\Big) \;\Big|\; (q, \gamma, \varepsilon, \beta, p) \in \Delta,\ q' \in Q\Big\}\\ A'' &:= A \times A''\\ \Gamma'' &:= \Gamma\end{aligned}$$

Beachte die Unterscheidung in der Definition von Δ'', ob es sich um einen ε-Übergang handelt. Diese ist notwendig, da δ' nur auf Zeichen $x \in \Sigma$ definiert ist!

(d) Nein ☺. Einfaches Gegenbeispiel: Schneide eine Typ-1 Sprache, die nicht Typ-2 ist, mit Σ^*. Etwas schwierigeres Gegenbeispiel: Schneide die Sprache L_1 mit $\mathrm{L}\big(a^* b^* a^* b^*\big)$.

Hausübung

Aufgabe H13 (Chomsky-Hierarchie) (4 Punkte)

Entscheiden und beweisen Sie für die folgenden Sprachen, ob sie kontextfrei sind oder nicht.

(a) $L_1 = \{a^n b^m \in \{a,b\}^* \mid n \geq m\}$

(b) $L_2 = \{a^p \in \{a\}^* \mid p \text{ prim}\}$

(c) $L_3 = \{a^{n!} \in \{a\}^* \mid n \geq 0\}$

Hinweis: Es gilt $(i+1)! - i > i!$ für alle $i \in \mathbb{N}$, $i > 1$ (Sie müssen diese Ungleichung nicht beweisen).

Lösung: Die Sprache L_1 ist kontextfrei, die Sprachen L_2 und L_3 sind das nicht.

(a) $\boxed{\text{1 P.}}$ Um zu zeigen, dass L_1 kontextfrei ist, können wir entweder eine kontextfreie Grammatik konstruieren oder einen Kellerautomaten angeben, der L_1 erkennt.

Eine kontextfreie Grammatik wäre:

$$X_0 \quad \to \quad aX_0 \mid aX_0 b \mid \varepsilon$$

Sei $\mathscr{P} = (\Sigma, Q, q_a, \Delta, A, \Gamma, \#)$ der Kellerautomat mit Eingabealphabet $\Sigma = \{a,b\}$, Zustandsmenge $Q = \{q_a, q_b\}$, $A = \{q_a, q_b\}$ als Menge der akzeptierenden Zustände, Kelleralphabet $\Gamma = \{\#, |\}$ und Übergangsrelation Δ gegeben durch

$$\begin{array}{llllll} \{ & (q_a, & \#, & \varepsilon, & \varepsilon, & q_a) \\ & (q_a, & \#, & a, & |, & q_a) \\ & (q_a, & |, & a, & ||, & q_a) \\ & (q_a, & |, & \varepsilon, & \varepsilon, & q_a) \\ & (q_a, & |, & b, & \varepsilon, & q_b) \\ & (q_b, & |, & b, & \varepsilon, & q_b) \quad \}. \end{array}$$

Dann erkennt $\mathscr{P}$ die Sprache L_1.

Um zu beweisen, dass L_2 und L_3 nicht kontextfrei sind, zeigen wir, dass in beiden Fällen das Pumping Lemma 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

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

(b) $\boxed{\text{1,5 P.}}$ Wir zeigen, dass L_2 nicht kontextfrei ist. Sei $i \in \mathbb{N}$ beliebig und betrachte

$$x = a^l,$$

wobei $l > i+1$ eine Primzahl ist. Offensichtlich ist $x \in L_2$ und $|x| \geq i$. Sei also $x = y \cdot u \cdot v \cdot w \cdot z$ eine Zerlegung von x mit $|uw| > 0$ und $|uvw| \leq i$. Aus $|uvw| \leq i$ folgt, dass $|yz| > 1$ und $|yvz| > 1$. Wählen wir $j = |yvz|$ und $x' = y \cdot u^j \cdot v \cdot w^j \cdot z$. Wir bestimmen die Länge von x':

$$|x'| = |y \cdot u^j \cdot v \cdot w^j \cdot z| = |yvz| + j|uw| = j(|uw| + 1).$$

Weil $j > 1$ und $|uw| + 1 > 1$, ist $|x'|$ nicht prim und $x' \notin L_2$.

(c) $\boxed{1{,}5\text{ P.}}$ Wir zeigen, dass L_3 nicht kontextfrei ist. Sei $i \in \mathbb{N}$ beliebig und betrachte

$$x = a^{(i+1)!}.$$

Offensichtlich ist $x \in L_2$ und $|x| \geq i$. Sei also $x = y \cdot u \cdot v \cdot w \cdot z$ eine Zerlegung von x mit $|uw| > 0$ und $|uvw| \leq i$. Für $i = 0$ haben wir dann einen Widerspruch ($0 < |uw| \leq |uvw| \leq 0$) und für $i = 1$ können wir $j = 2$ wählen (für $i = 1$ und $j = 2$ ist $x = aa$, $uw = a$ und $y \cdot u^j \cdot v \cdot w^j \cdot z = a^3 \notin L_2$). Wir nehmen deshalb an, dass $i > 1$, und verwenden, dass dann $(i+1)! - i > i!$. Jetzt wählen wir $j = 0$ und behaupten

$$x' = y \cdot u^j \cdot v \cdot w^j \cdot z = yvz \notin L_2.$$

Da $i \geq |uw| > 0$, kann die Länge von x' unmöglich die Form $n!$ haben:

$$(i+1)! = |x| > |x| - |uw| = |x'| \geq (i+1)! - i > i!.$$

Aufgabe H14 (Grammatik einer Programmiersprache) (2 Punkte)
Finden Sie die Programmiersprache BF auf der deutschen Wikipedia. Geben Sie eine kontextfreie Grammatik für ihre Syntax an.

Lösung: Eine mögliche kontextfreie Grammatik ist $G = \big(\Sigma = \{\texttt{>}, \texttt{<}, \texttt{+}, \texttt{-}, \texttt{.}, \texttt{,}, \texttt{[}, \texttt{]}\}, V = \{X\}, P, X\big)$ mit den folgenden Produktionen:

$$X \rightarrow \texttt{>}X \mid \texttt{<}X \mid \texttt{+}X \mid \texttt{-}X \mid \texttt{.}X \mid \texttt{,}X \mid X\,\texttt{[}X\texttt{]}X \mid \varepsilon$$

Aufgabe H15 (Entscheidbarkeit des Wortproblems) (4 Punkte)
Sei $\Sigma = \{a, b\}$. Eine *Instanz* des *Post'schen Korrespondenzproblems (PKP)* ist eine endliche Folge von Wortpaaren $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$ mit $x_i, y_i \in \Sigma^+$. Gibt es nun eine Folge von Indizes $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, k\}$ $(n \geq 1)$ mit

$$x_{i_1} x_{i_2} \ldots x_{i_n} = y_{i_1} y_{i_2} \ldots y_{i_n},$$

so heißt diese *Lösung* der PKP-Instanz $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$.

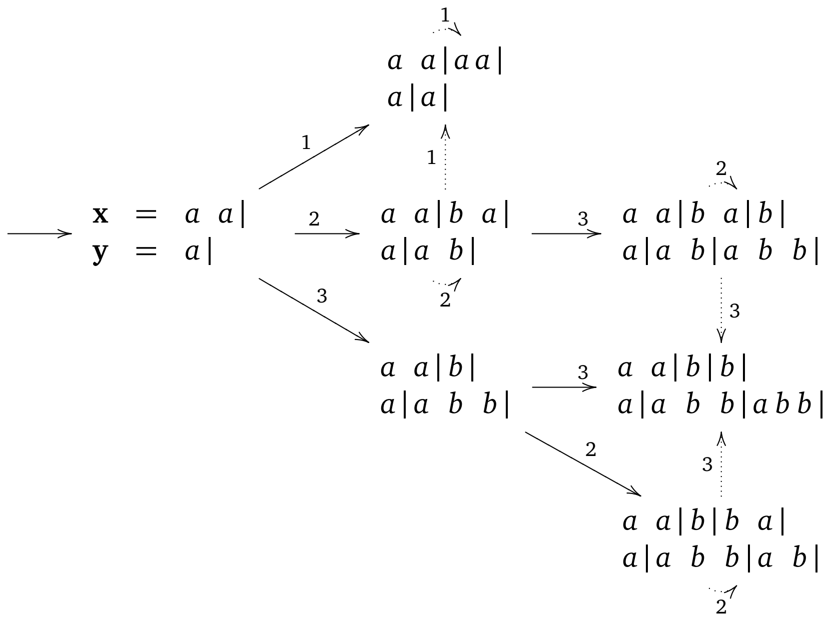
(a) Entscheiden Sie für die folgenden Instanzen der PKP, ob es eine Lösung gibt. Falls es eine Lösung gibt, geben Sie diese an. Falls nicht, beweisen Sie, dass es keine Lösung gibt.
   i. $(abba, a), (aa, bbaa), (bbbb, abb)$
   ii. $(aa, a), (ba, ab), (b, abb)$

(b) Geben Sie eine Grammatik an, die genau alle endlichen Folgen von Wortpaaren $(x_1, y_1), \ldots, (x_k, y_k)$ mit $x_i, y_i \in \Sigma^+$ (d. h. die syntaktisch korrekten Instanzen des PKP) erzeugt.

Lösung:

(a) i. $\boxed{1\text{ P.}}$ Eine Lösung der Instanz ist $(1, 2, 3, 2)$.

   ii. $\boxed{2\text{ P.}}$ Angenommen es gibt eine Lösung $(i_1, \ldots, i_n)$ mit $\mathbf{x} = \mathbf{y}$, wobei $\mathbf{x} = x_{i_1} \ldots x_{i_n}$ und $\mathbf{y} = y_{i_1} \ldots y_{i_n}$. Da beide Worte mit demselben Buchstaben beginnen müssen, ist $i_1 = 1$. Dem formalen Beweis stellen wir zunächst das nachfolgende Transitionssystem voran, das alle möglichen Erweiterungen einer potentiellen Lösung um je ein Tupel der Instanz beschreibt. Gestrichelte Pfeile deuten an, dass mit dem angegebenen Tupel eine äquivalente Konstellation auftritt. Gibt es keinen ausgehenden Pfeil mehr, so lässt sich die jeweilige Folge nicht mehr fortsetzen, da **x** und **y** sich für jedes mögliche angehangene Tupel in mindestens einer Position unterscheiden würden.



Ist $i_2 = 1$, so kann nur noch $i_k = 1$ sein für $k \geq 3$, was jedoch nie zu einer Lösung führt.
Ist $i_2 = 2$, so muss y_{i_3} ebenfalls mit a beginnen und es folgt wiederum $i_3 \in \{2, 3\}$. Stets $i_k = 2$ für alle $k \geq 2$ zu wählen führt zu *keiner* endlichen Lösung der Instanz. Entsprechend (da die Existenz einer Lösung vorausgesetzt wurde) muss es ein i_k mit $k \geq 2$ geben, so dass $i_k = 3$ gewählt wurde. Im Falle $k = 2$ war $i_1 = 1$, womit nun x_{i_3} mit b beginnen muss. Wird nun $i_3 = 3$ gewählt, so gibt es fortan keine Fortsetzung mehr, so dass die Instanz eine Lösung besitzt, denn es müsste x_{i_4} mit abb beginnen, was nicht möglich ist.
Im Falle $k > 2$ muss $x_{i_{k+1}}$ mit b beginnen. Wird $i_{k+1} = 2$ gewählt, so kann wieder unendlich fortgesetzt werden, ohne eine endliche Lösung zu erhalten. Wird hingegen $i_{k+1} = 3$ gewählt, so müsste wiederum $x_{i_{k+2}}$ mit abb beginnen, was nicht möglich ist. Entsprechend hat diese Instanz *keine* Lösung.

(b) $\boxed{\text{1 P.}}$ Eine mögliche Grammatik ist $G = (\Sigma, V, P, S)$ mit $V = \{S, T, U\}$ und den Produktionen

$$\begin{aligned} S &\to T\#S \mid T \\ T &\to U\$U \\ U &\to a_iU \mid a_i \quad \text{(für alle } a_i \in \Sigma) \end{aligned}$$

Minitest

Aufgabe M16 (Chomsky-Hierarchie)
Welche der folgenden Aussagen ist korrekt, welche nicht?

- ☐ Sind L_1 und L_2 kontextfreie Sprachen, dann ist auch $L_1 \setminus L_2$ kontextfrei.
- ☐ Sind L_1 und L_2 kontextfreie Sprachen, dann ist $L_1 \setminus L_2$ kontextsensitiv.
- ☐ Jede Sprache mit endlichem Komplement ist regulär.
- ☐ Jede Sprache mit endlichem Komplement ist kontextfrei.
- ☐ Die Sprache aller regulären Ausdrücke ist regulär.
- ☐ Die Sprache aller regulären Ausdrücke ist kontextfrei.

Lösung:

- ☐ Wenn die Aussage gelten würde, wären kontextfreie Sprachen unter Komplement abgeschlossen (nimm $L_1 = \Sigma^*$), sind sie aber nicht.
- ☒ Die Klasse der kontextfreien Sprachen ist in der Klasse der kontextsensitiven enthalten und letztere ist abgeschlossen unter Schnitt und Komplementbildung.
- ☒ Endliche Sprachen sind regulär und reguläre Sprachen sind unter Komplement abgeschlossen.
- ☒ Aus der vorherigen Frage wissen wir bereits um die Regularität, daher ist die Sprache insbesondere kontextfrei.
- ☐ Die Klammern lassen sich nicht regulär abbilden.
- ☒ Aus der rekursiven Definition regulärer Ausdrücke lässt sich ablesen, wie eine kontextfreie Grammatik für die *Syntax* eines regulären Ausdrucks auszusehen hat.

Aufgabe M17 (Kellerautomaten)
Sei $\mathscr{P} = \left(\Sigma = \{a, b\}, Q = \{q_0\}, q_0, \Delta, A = \{q_0\}, \Gamma = \{\#\}, \#\right)$ ein Kellerautomat mit

$$\Delta = \{(q_0, \#, a, \#\#, q_0), (q_0, \#, b, \varepsilon, q_0)\}.$$

(a) Welche der folgenden Worte werden von $\mathscr{P}$ erkannt?

- ☐ b
- ☐ ab
- ☐ ba
- ☐ abb
- ☐ $abba$
- ☐ $ababababa$

(b) Was ist der Typ der von $\mathscr{P}$ erkannten Sprache? (Kreuzen Sie *alle* zutreffenden Antworten an.)
☐ 3 ☐ 2 ☐ 1 ☐ 0

Lösung:

(a)
- ☒ b
- ☐ ab
- ☐ ba
- ☒ abb
- ☐ $abba$
- ☐ $ababababa$

(b) ☐ 3 ☒ 2 ☒ 1 ☒ 0

Sei L die von $\mathscr{P}$ erkannte Sprache. Da L durch einen Kellerautomaten gegeben ist, ist sie kontextfrei (und somit vom Typ 2, 1, 0). Die Frage ist nur noch, ob L auch regulär (Typ 3) ist.

Die Sprache L enthält die Wörter w mit $|w|_a + 1 = |w|_b$ (also die Anzahl von b ist für eins größer als die Anzahl von a) und für alle Präfixe $p \neq w$ von w muss gelten $|p|_a \geq |p|_b$. Man zeigt, dass L nicht regulär ist, z. B. mithilfe des Pumping-Lemmas durch Betrachten der Worte der Form $a^n b^{n+1}$.