Universität Siegen
Lehrstuhl Theoretische Informatik
Markus Lohrey

Übungsblatt 12

Aufgabe 1. Geben Sie einen Kellerautomaten und eine kontextfreie Grammatik an, die die folgende Sprache akzeptiert:

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

Lösung zu Aufgabe 1. Grammatik

- $G = (V, \Sigma, P, S)$
- $V = \{S, A, B\}$
- $\Sigma = \{a, b\}$
- $P = \{S \to aSb \mid A \mid B,\ A \to aA \mid a,\ B \to bB \mid b\}$

Über die Regel $S \to aSb$ erzeugen wir Wörter mit gleich vielen a's und b's.

Die Übergänge $S \to A$ und $S \to B$ fügen in der Mitte dann noch ein Wort a^n oder b^n mit $n \geq 1$ ein, so dass am Ende entweder mehr a's oder mehr b's erzeugt wurden.

Kellerautomat

- $M = (Z, \Sigma, \Gamma, \delta, z_0, \#)$
- $Z = \{z_0, z_1, z_2, z_3, s_0, s_1, q_0\}$
- $\Sigma = \{a, b\}$
- $\Gamma = \{\#, A\}$

Dabei ist δ wie folgt definiert:

$$
\begin{aligned}
\delta : (z_0, a, \#) &\to (z_1, \#) && (1)\\
(z_1, a, \#) &\to (z_1, \#) && (1)\\
(z_1, \varepsilon, \#) &\to (z_1, \varepsilon) && (1)\\
\\
(z_0, b, \#) &\to (z_2, \#) && (2)\\
(z_2, b, \#) &\to (z_2, \#) && (2/4)\\
(z_2, \varepsilon, \#) &\to (z_2, \varepsilon) && (2/4)\\
\\
(z_0, a, \#) &\to (z_3, A\#) && (3/4)\\
(z_3, a, A) &\to (z_3, AA) && (3/4)\\
(z_3, b, A) &\to (s_0, \varepsilon) && (3)\\
(s_0, b, A) &\to (s_0, \varepsilon) && (3)\\
(s_0, \varepsilon, A) &\to (s_1, \varepsilon) && (3)\\
(s_1, \varepsilon, A) &\to (s_1, \varepsilon) && (3)\\
(s_1, \varepsilon, \#) &\to (s_1, \varepsilon) && (3)\\
\\
(z_3, b, A) &\to (q_0, \varepsilon) && (4)\\
(q_0, b, A) &\to (q_0, \varepsilon) && (4)\\
(q_0, b, \#) &\to (z_2, \#) && (4)
\end{aligned}
$$

Lassen Sie sich nicht abschrecken von den vielen Transitionen, denn das Prinzip ist einfacher als es aussieht. Wir unterscheiden nichtdeterministisch 4 Typen von Wörtern, die zur Sprache gehören:

(1) a^n für $n \geq 1$

(2) b^n für $n \geq 1$

(3) $a^m b^n$ mit $m, n \geq 1$ und $m > n$

(4) $a^m b^n$ mit $m, n \geq 1$ und $m < n$

Die ersten beiden Fälle (1) und (2) sind einfach, da entweder nur a's oder b's gelesen werden und anschließend kann man das Kellerbodensymbol entfernen.
In den Fällen (3) und (4) wird zu Beginn für jedes gelesene a ein A auf den

Keller gelegt bis das erste b gelesen wird. Anschließend wird im Fall (3) für jedes b ein A entfernt, und dann wechselt man mit einem ε-Übergang in den Zustand s_1 falls weiterhin ein A auf dem Keller liegt (und somit mehr a's als b's gelesen wurden). Der Zustand s_1 leert anschließend den Keller ohne weitere Buchstaben zu lesen. Beachten Sie, dass falls man den ε-Übergang anwendet bevor das Wort zu Ende gelesen wurde, so wird am Ende <u>nicht</u> akzeptiert, da im Zustand s_1 keine weiteren Buchstaben gelesen werden.
Im Fall (4) wird ebenfalls für jedes b ein A vom Keller entfernt, aber anschließend wechselt man in den Zustand z_2, falls man das Kellerbodensymbol # erreicht und ein b gelesen wird (und somit mehr b's als a's gelesen wurden). Im Zustand z_2 werden dann noch beliebig viele b's gelesen bevor man das Kellerbodensymbol entfernt.

Aufgabe 2. Betrachten Sie noch einmal Aufgabe 2 (c) von Blatt 11. Eine kontextfreie Grammatik, die die Sprache $L = \{w \in \{a,b\}^* \mid w = w^r\}$ erzeugt, ist $G = (V, \Sigma, P, S)$ mit $V = \{S\}$, $\Sigma = \{a, b\}$ und

$$P = \{S \to \varepsilon \mid a \mid b \mid aSa \mid bSb\}.$$

Wandeln Sie die Grammatik mit dem Verfahren der Vorlesung (Folie 279) in einen PDA um. Überprüfen Sie, ob ihr Automat richtig arbeitet, indem Sie ihn auf den Eingaben

- $w_1 = abbabba$ und
- $w_2 = abb$

laufen lassen.

Lösung zu Aufgabe 2. Nach dem Verfahren der Vorlesung erhalten wir einen Kellerautomaten $M = (\{z\}, \Sigma, V \cup \Sigma, \delta, z, S)$ mit δ gegeben durch

$$\begin{aligned}
\delta(z, \varepsilon, S) &= \{(z, \varepsilon), (z, a), (z, b), (z, aSa), (z, bSb)\},\\
\delta(z, a, a) &= \{(z, \varepsilon)\},\\
\delta(z, b, b) &= \{(z, \varepsilon)\}.
\end{aligned}$$

Beachte: Dieser Automat ist deutlich kleiner als der Automat aus der Lösung zu Aufgabe 2 (c). Allerdings arbeitet der PDA weniger intuitiv. Dies liegt daran, dass im Keller des Automaten die Ableitung durch die Grammatik G simuliert wird, wohingegen der Automat aus der Lösung des vorigen Blattes nur eine „Stackphase“, eine „Übergangsphase“ und eine „Löschphase“ hat. Man beachte ferner, dass bei der neuen Konstruktion das Kelleralphabet alle Terminalsymbole enthält. Generell sind Kelleralphabete durch diese Konstruktion größer.

Wir sehen sofort, dass w_1 ein Palindrom ist. Wir haben:

$$\begin{aligned}(z, abbabba, S) &\vdash (z, abbabba, aSa) \vdash (z, bbabba, Sa) \vdash (z, bbabba, bSba)\\ &\vdash (z, babba, Sba) \vdash (z, babba, bSbba) \vdash (z, abba, Sbba)\\ &\vdash (z, abba, abba) \vdash^* (z, \varepsilon, \varepsilon)\end{aligned}$$

Also wird w_1 von M akzeptiert. Für w_2 bekommen wir zum Beispiel

$$(z, abb, S) \vdash (z, abb, aSa) \vdash (z, bb, Sa) \vdash (z, bb, ba) \vdash (z, b, a)$$

Wegen des Nichtdeterminismus müssen wir uns noch überlegen, dass es nicht doch eine Möglichkeit gibt $(z, \varepsilon, \varepsilon)$ zu erhalten: Wir müssen das Eingabewort komplett abarbeiten, daher müssen wir im ersten Schritt einen Übergang nehmen, bei dem S erhalten bleibt und dafür kommt (wegen erstes Zeichen a) nur einer in Frage ($S \to aSa$). Das Wort w_2 hat Länge 3 und wir wissen daher, dass im nächsten Schritt die Mitte gelesen werden muss. Es bleibt daher nur den zu $S \to b$ korrespondierenden Übergang zu nehmen. Obwohl wir bis hierher also stets die sinnvollsten Übergange gewählt haben, bleibt der Automat nun bei (z, b, a) stehen.

Es gäbe zum Beispiel auch die Möglichkeit das Eingabewort komplett abzuarbeiten, wodurch wir (z, ε, bba) erhalten. Hier ist allerdings der Keller nicht leer.

Aufgabe 3. Sind die folgenden Aussagen wahr oder falsch?

(a) Wenn L eine Sprache mit $\text{index}(R_L) = \infty$ ist, dann ist L kontextfrei.

(b) Wenn L eine nicht kontextfreie Sprache ist, dann ist $\text{index}(R_L) = \infty$.

(c) Wenn G eine Grammatik in CNF ist, dann ist $L(G)$ kontextfrei und nicht regulär.

(d) Es existieren kontextfreie Sprachen L_1 und L_2 so, dass $L_1 \cap L_2$ auch kontextfrei ist.

(e) Für jede Sprache L mit $L = L_1 \cap L_2$, wobei L_1 nicht regulär und L_2 nicht kontextfrei ist, ist L nicht regulär.

(f) Für jedes nicht unäre Alphabet Σ existieren unendlich viele kontextfreie, nicht reguläre Sprachen $L_i \subseteq \Sigma^*$ mit folgender Eigenschaft: Sei $\mathcal{H}$ die Menge aller Homomorphismen $h : \Sigma^* \to \Sigma^*$. Dann ist $\bigcup_{h \in \mathcal{H}} h(L_i)$ regulär.

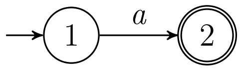
Lösung zu Aufgabe 3.

(a) falsch, weil für alle nicht-regulären Sprachen L auch $\text{index}(R_L) = \infty$ gilt, aber nicht jede nicht-reguläre Sprache ist eine kontextfreie Sprache, zum Beispiel $L = \{a^n b^n c^n \mid n \in \mathbb{N}\}$.

(b) wahr, weil jede Sprache, die nicht kontextfrei ist, auch nicht regulär ist und somit $\text{index}(R_L) = \infty$ gilt.

(c) falsch
Sei z.B. $G = (\{S\}, \{a\}, P, S)$ mit $P = \{S \to a\}$.

Diese Grammatik ist in CNF, aber es gilt $L(G) = \{a\}$ und das ist eine reguläre Sprache. Sie wird beispielsweise vom nachfolgenden NFA erzeugt:

M_G: 

(d) wahr
Sei z.B. $L_1 = L_2$, dann ist $L_1 \cap L_2 = L_1 = L_2$ und somit genau dann kontextfrei, wenn L_1 und L_2 kontextfrei sind.

(e) falsch
Die Sprache $L_2 = \{ww \mid w \in \{a, b\}^*\}$ ist nicht kontextfrei und $L_1 = \overline{L_2}$ ist nicht regulär (siehe Aufgabe 4 von Blatt 9). Es gilt aber $L_1 \cap L_2 = \varnothing$, welches eine reguläre Sprache ist.

(f) wahr
Sei $a \in \Sigma$. Es gibt unendlich viele nicht reguläre, aber kontextfreie Sprachen mit $a \in L_i$, zum Beispiel $L_i = \{a^n b^n \mid n \geq i\} \cup \{a\}$ für $\Sigma = \{a, b\}$. Die Menge aller Homomorphismen $\mathcal{H}$ enthält insbesondere folgende Abbildungen: $h_w : \Sigma^* \to \Sigma^*$ mit $h_w(a) = w$ und $h_w(s) = \varepsilon$ mit $w \in \Sigma^*$ und $s \in \Sigma \backslash \{a\}$. Also enthält $h_w(L_i)$ das Wort w und $\bigcup_{h \in \mathcal{H}} h(L_i)$ enthält Σ^*. Somit haben wir für jedes $i \in \mathbb{N}$ eine Sprache L_i mit $\bigcup_{h \in \mathcal{H}} h(L_i) = \Sigma^*$ (und Σ^* ist regulär).