Institut für Theoretische Informatik
Lehrstuhl Prof. Dr. D. Wagner

Übungsblatt 7

Vorlesung Theoretische Grundlagen der Informatik im WS 19/20

Ausgabe: 21. Januar 2020
Abgabe: 4. Februar 2020, 11:00 Uhr (im Kasten im UG von Gebäude 50.34)

Aufgabe 1

(3 + 1 + 1 = 5 Punkte)

Sei $G = (\Sigma, V, S, R)$ mit $\Sigma = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}\}$ und $V = \{S, A, B, C, D, E, F\}$ die durch folgende Regelmenge gegebene Grammatik:

$$\begin{aligned}
S &\rightarrow C \mid FA \mid D \\
A &\rightarrow F\mathsf{c}A \mid \mathsf{d} \mid C\mathsf{b} \\
B &\rightarrow \mathsf{a} \mid C\mathsf{b}A \mid \mathsf{a}AE \\
C &\rightarrow \mathsf{c}DA \mid \mathsf{a}A \mid F\mathsf{cb} \\
D &\rightarrow \mathsf{c}F\mathsf{d} \mid FDA \\
E &\rightarrow \mathsf{a}F\mathsf{a} \mid CE\mathsf{d} \mid \mathsf{a} \\
F &\rightarrow \mathsf{b}D \mid \mathsf{b}FC
\end{aligned}$$

(a) Identifizieren Sie alle nutzlosen Variablen in G mit dem Verfahren aus der Vorlesung. Geben Sie die Grammatik G' an, die durch Entfernen der nutzlosen Variablen entsteht.

(b) Ist G' minimal in dem Sinne, dass es keine Grammatik mit weniger Variablen gibt, die $L(G')$ erzeugt? Begründen Sie Ihre Antwort.

(c) Ist $L(G)$ endlich? Begründen Sie Ihre Antwort.

Hinweis: Die Grammatik muss dazu nicht notwendigerweise in Chomsky-Normalform gebracht werden.

Aufgabe 2

(3 Punkte)

Sei $L = \{x \,\texttt{\#}\, y^{\mathsf{R}} \mid x, y \text{ sind binär kodierte Zahlen mit } x + 1 = y\}$. Geben Sie formal einen deterministischen Kellerautomaten an, der L erkennt.

In einer früheren Version des Übungsblatts wurde gefordert, dass x und y keine führenden Nullen enthalten. Diese Forderung wurde gestrichen. Für die volle Punktzahl reicht es, wenn Ihr Automat auf solchen Eingaben korrekt arbeitet, in denen keine führenden Nullen vorkommen. Eingaben mit führenden Nullen müssen nicht korrekt behandelt werden.

Aufgabe 3 (2 + 2 + (2) + 2 + 1 = 7 + (2) Punkte)

Aus der Vorlesung wissen Sie, dass eine Linksableitung eine Ableitung ist, bei der in jedem Schritt das linkeste Nichtterminalsymbol des aktuellen Wortes aus $(\Sigma \cup V)^*$ abgeleitet wird. Sei $G = (\Sigma, V, S, R)$ eine kontextfreie Grammatik. Zu jedem Wort $w \in L(G)$ gibt es eine Linksableitung bezüglich G.

(a) Geben Sie einen nichtdeterministischen Kellerautomaten an, der $L(G)$ erkennt.

Anmerkung: Konstruieren Sie keine zu G äquivalenten Grammatiken, sondern denken Sie an das Konzept der Linksableitung.

(b) Beweisen Sie, dass Ihr Kellerautomat genau $L(G)$ erkennt.

In Anbetracht von (a) und (b) stellt man sich die Frage, wozu die Greibach-Normalform überhaupt nützlich ist. Das TGI-Team hat leider keine überzeugende Antwort. Und Sie?

(c) *Bonusaufgabe:* Erklären Sie möglichst überzeugend, was der Nutzen der Greibach-Normalform ist.

Diese Bonusaufgabe wird mit maximal 2 Punkten bewertet.

Wir versuchen nun, den Begriff der Linksableitung sinnvoll für kontextsensitive Sprachen zu erweitern: Eine *Präfixableitung* ist eine Ableitung, bei der in jedem Schritt ein Präfix des aktuellen Wortes aus $(\Sigma \cup V)^*$ abgeleitet wird. Eine Linksableitung ist also ein Spezialfall der Präfixableitung, bei der jedes Präfix Länge Eins hat.

(d) Sei G eine kontextsensitive Grammatik, sodass zu jedem Wort $w \in L(G)$ eine Präfixableitung existiert. Erklären Sie, wie Sie Ihren Ansatz aus Teilaufgabe (a) erweitern können, um einen Kellerautomaten zu konstruieren, der $L(G)$ erkennt.

(e) Geben Sie eine kontextsensitive Grammatik G und ein Wort $w \in L(G)$ an, sodass w nicht durch eine Präfixableitung abgeleitet werden kann.

Aufgabe 4 (1 + 2 + 2 + 1 = 6 Punkte)

In Vorlesung und Übung wurden bisher *nichtdeterministische* Kellerautomaten betrachtet. Sei $\mathcal{A}$ nun ein *deterministischer* Kellerautomat, d.h. es gibt keine ε-Übergänge und für jedes Eingabesymbol ist der Übergang von $\mathcal{A}$ eindeutig definiert. Gehen Sie davon aus, dass das Eingabealphabet Σ mindestens zwei Symbole enthält, dass $\mathcal{A}$ durch akzeptierende Endzustände akzeptiert und dass der Stack von $\mathcal{A}$ niemals leer ist.

Definiere für solche deterministischen Kellerautomaten das Konzept der *Essenz* folgendermaßen. Für jedes Wort w existiert ein[1] Wort w', sodass die Höhe des Stacks nach Verarbeitung von ww' minimal unter allen möglichen w' ist. Ist nach Abarbeitung von ww' die Höhe des Stacks h, so ist also für jedes Wort v die Höhe des Stacks nach Abarbeitung von $ww'v$ mindestens h. Sei $\alpha_1\alpha_2 \ldots \alpha_h$ der Inhalt des Stacks nach Abarbeitung von ww'. Das oberste Symbol auf dem Stack ist $\hat{\alpha} := \alpha_h$. Der aktuelle Zustand sei q. Bezeichne $(q, \hat{\alpha})$ als die *Essenz* von w.

[1] Das Wort w' ist nicht notwendigerweise eindeutig.

(a) Zeigen Sie, dass es mindestens zwei unterschiedliche Wörter $w_1 \neq w_2$ mit der gleichen Länge und derselben Essenz gibt.

Hinweis: Wie viele unterschiedliche Essenzen kann es höchstens geben?

(b) Seien w_1, w_2 Wörter mit derselben Essenz $(q, \hat{\alpha})$. Zeigen Sie, dass dann für jedes Wort $v \in \Sigma^\star$ gilt:

$$w_1 w_1' v \in L(\mathcal{A}) \iff w_2 w_2' v \in L(\mathcal{A})$$

Hinweis: Erinnern Sie sich daran, dass die Konfiguration (q, v, α) eines Kellerautomaten aus drei Teilen besteht: dem aktuellen Zustand q, dem Teil der Eingabe $v \in \Sigma^\star$, der noch nicht gelesen wurde, und dem Stackinhalt $\alpha \in \Gamma^\star$.

(c) Definieren Sie die Sprache, die genau aus allen Palindromen besteht. Nutzen Sie das Ergebnis der letzten Teilaufgabe, um zu zeigen, dass $\mathcal{A}$ die Palindromsprache nicht erkennen kann.

(d) Bestimmen Sie das minimale k, sodass deterministische Kellerautomaten alle Sprachen von Typ-k erkennen können. Begründen Sie kurz.

Aufgabe 5 (1 + 2 + 3 + 1 = 7 Punkte)

Sie und Ihre Kommilitonen Lena und Leopold haben jeweils einen nichtdeterministischen Kellerautomaten erworben. Lena klaut nun den Stack von Leopolds Kellerautomat und baut ihn in ihrem eigenen Kellerautomat als Zweitstack ein, den sie unabhängig vom Erststack benutzen kann. Wir betrachten nun also drei verschiedene Maschinenmodelle: Leopolds Kellerautomat NPDA_0 hat überhaupt keinen Stack mehr, Ihr Kellerautomat NPDA_1 ist ein ganz normaler Kellerautomat mit einem Stack, und Lenas Kellerautomat NPDA_2 hat zwei Stacks.

(a) Leopold fragt sich, ob sein Kellerautomat NPDA_0 jetzt weniger Sprachen erkennen kann als Ihr Kellerautomat NPDA_1. Beantworten Sie diese Frage, indem Sie begründen, welche Klasse von Sprachen von NPDA_0 erkannt werden können.

Lena hofft, dass sie mit ihrem Kellerautomat NPDA_2 mehr Sprachen erkennen kann als Ihr Kellerautomat NPDA_1.

(b) Zeigen Sie, dass Lena recht hat, indem Sie eine Sprache angeben, die von NPDA_2 erkannt wird, von NPDA_1 aber nicht erkannt werden kann. Begründen Sie beides!

(c) Welche Klasse von Sprachen kann NPDA_2 erkennen?
Hinweis: Kann NPDA_2 ein mächtigeres Berechnungsmodell simulieren?

(d) Kann Lena ihren Kellerautomaten noch mächtiger machen, indem sie einen dritten Stack einbaut? Begründen Sie!

Aufgabe 6 (1 + 1 + 1 = 3 Punkte)

Hinweis: Die für diese Aufgabe notwendigen Grundlagen zur Huffman-Kodierung werden in der Vorlesung vom 28.1. vorgestellt.

Die folgende Tabelle gibt eine Häufigkeitsverteilung für die Zeichen des Alphabets $\Sigma = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}, \mathsf{e}, \mathsf{f}, \mathsf{g}, \mathsf{h}\}$ an.

a	b	c	d	e	f	g	h
17%	5%	12%	23%	8%	11%	18%	6%

(a) Nehmen Sie an, Sie möchten Wörter über dem Alphabet Σ binär kodieren. Was ist die erwartete Länge der Kodierung eines Wortes in Abhängigkeit von dessen Länge n, wenn Sie die obige Verteilung und folgende naive Binärkodierung der Zeichen zugrunde legen?

a	b	c	d	e	f	g	h
000	001	010	011	100	101	110	111

(b) Konstruieren Sie den Huffman-Kodierungsbaum zu der obigen Häufigkeitsverteilung.

(c) Was ist die erwartete Länge der Kodierung eines Wortes in Abhängigkeit von dessen Länge n, wenn sie die obige Verteilung und Ihre Huffman-Kodierung zugrunde legen?