Theoretische Informatik 2

Sommersemester 2017

Georg Schnitger
Philipp Tertel
Hung The Tran

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Institut für Informatik
AG Theoretische Informatik

Übung 4

Ausgabe: 16.05.17
Abgabe: 23.05.17

Auf diesem Blatt können bis zu sechs Extrapunkte erreicht werden: Nur 24 der 30 möglichen Punkte werden in der Endabrechnung aller möglichen Punkte erfasst.

Aufgabe 4.1. ***Eindeutigkeit – Mehrdeutigkeit*** (4+2+4 Punkte)

a) Die Dyck-Sprache D aller wohlgeformten Klammerausdrücke wird sowohl definiert durch die kontextfreie Grammatik G_1 mit den Produktionen

$$S_1 \to (S_1) \mid S_1 S_1 \mid \varepsilon$$

wie auch durch die kontextfreie Grammatik G_2 mit den Produktionen

$$S_2 \to (S_2)S_2 \mid \varepsilon$$

Welche der Grammatiken ist eindeutig, welche ist mehrdeutig? Beweisen Sie ihre Antwort. (Sie brauchen nicht zu zeigen, dass die jeweilige Grammatik die Dyck-Sprache erzeugt.)

b) Die Grammatik G mit dem Terminalalphabet $\Sigma := \{\mathtt{i}, \mathtt{e}\}$ und den Produktionen

$$S \to \mathtt{i}\ S\ \mathtt{e}\ S \mid\ \mathtt{i}\ S \mid \varepsilon$$

modelliert eine Spezifikation der <u>i</u>f-then-<u>e</u>lse-Anweisung. Leider besitzt das Wort

$$\mathtt{i\ i\ e}$$

zwei verschiedene Ableitungsbäume. Das Ziel dieser Aufgabe ist der Entwurf einer eindeutigen kontextfreien Grammatiken für $L(G)$.

i) Wir betrachten die Grammatik G_1 mit den Produktionen

$$S \to S\ \mathtt{i}\ \mid\ \mathtt{i}\ S\ \mathtt{e}\ \mid \varepsilon$$

Zeigen Sie: G_1 ist eindeutig.

ii) Leider gilt nicht $L(G) = L(G_1)$. Entwerfen Sie eine eindeutige Grammatik G_2 mit $L(G) = L(G_2)$. Ihre Grammatik sollte einem „`e`“ das nächstliegende linke, offene „`i`“ zuweisen. Was bedeutet das? Für eine Produktion der Form „$Z \to \mathtt{i}\ X\ \mathtt{e}\ Y$“ ist zu garantieren, dass X nur if-then-else-Ausdrücke mit derselben Anzahl von `i`’s und `e`’s erzeugt.

Eine *informelle* Begründung, warum Ihre Konstruktion funktioniert, ist völlig ausreichend.

Bitte wenden!

Aufgabe 4.2. ***Kontextfreie Sprachen auf unären Alphabeten*** (4+4+2 Punkte)

Sei $\Sigma = \{1\}$ und sei $L \subseteq \Sigma^*$ eine kontextfreie Sprache. Wir möchten mit Hilfe des Pumping-Lemmas für kontextfreie Sprachen zeigen, dass L regulär ist.

a) Sei N die Pumpingkonstante, wenn das Pumping-Lemma für L angewandt wird. Zeigen Sie: Für jedes $M \geq N$ mit $1^M \in L$ gibt es eine Zahl $1 \leq a_M \leq N$, so dass

$$\{\, 1^{M-a_M} 1^{i \cdot a_M} \;:\; i \in \mathbb{N} \,\} \subseteq L.$$

b) Zeigen Sie, dass $L' := \{1^M \in L \;:\; M \geq N\}$ eine *endliche* Vereinigung von Mengen der Form $\{1^b 1^{i\cdot a} \;:\; i \in \mathbb{N}_{>0}\}$ für Zahlen $a, b \in \mathbb{N}$ mit $a \leq N$ und $b \geq N - a$ ist.

c) Zeigen Sie, dass L regulär ist.

Kommentar: L ist also die Vereinigung von „linearen“ Sprachen.

Aufgabe 4.3. ***Inhärente Mehrdeutigkeit*** (3+4+3 Punkte)

Die Sprache

$$L = \{\, a^i b^j c^k \;:\; i = j \text{ oder } j = k \,\}$$

ist kontextfrei. Es soll gezeigt werden, dass L inhärent mehrdeutig ist, dass L also von keiner eindeutigen, kontextfreien Grammatik erzeugt werden kann. Wenden Sie dazu Ogden's Lemma auf eine vermeintlich eindeutige Grammatik G für L an. Sei N die Pumping-Konstante.

a) Wählen Sie das Wort $z_1 := a^M b^M c^{M+M!} \in L$ (mit $M := \max\{N, 3\}$) und markieren Sie alle a's. Aus dem Beweis von Ogden's Lemma erhalten Sie

- eine Zerlegung $z_1 = u_1 v_1 w_1 x_1 y_1$ (mit mindestens einer markierten Position in $v_1 x_1$ und höchstens N markierten Positionen in $v_1 w_1 x_1$) sowie
- eine Variable A mit den Ableitungen $S \overset{*}{\Rightarrow} u_1 A y_1$ sowie $A \overset{*}{\Rightarrow} v_1 A x_1 \mid w_1$.

Wir schauen noch etwas genauer hin.

i) Zeigen Sie: $v_1 = a^j$ und $x_1 = b^j$ für ein j mit $1 \leq j \leq M$. In der Zerlegung $z_1 := u_1 v_1 w_1 x_1 y_1$ treten c's also nur in y_1 auf.

ii) Zeigen Sie: Das Wort

$$z^* := a^{M+M!} b^{M+M!} c^{M+M!}$$

kann allein mit den Ableitungen $S \overset{*}{\Rightarrow} u_1 A y_1$ und $A \overset{*}{\Rightarrow} v_1 A x_1 \mid w_1$ abgeleitet werden. Es werden mehr als $M!$ viele b's durch die Ableitung $A \overset{*}{\Rightarrow} v_1 A x_1$ eingeführt.

b) Wenn Sie das Vorgehen aus a) auf das Wort $z_2 := a^{M+M!} b^M c^M$ (diesmal mit markierten c's) anwenden, erhalten Sie eine Zerlegung $z_2 = u_2 v_2 w_2 x_2 y_2$ mit $v_2 = b^k, x_2 = c^k$ (für $1 \leq k \leq M$) sowie eine Variable B mit den Ableitungen $S \overset{*}{\Rightarrow} u_2 B y_2$ und $B \overset{*}{\Rightarrow} v_2 B x_2 \mid w_2$. Auch jetzt kann man das Wort z^* mit diesen neuen Ableitungen erzeugen.

Zeigen Sie: Es gibt keinen Ableitungsbaum, der beide Ableitungen von z^* ausdrückt.

Fazit: Damit haben wir die Mehrdeutigkeit der Sprache L gezeigt.