Theoretische Informatik 2

Sommersemester 2018

Elizaveta Kovalevskaya
Georg Schnitger
Philipp Tertel

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Institut für Informatik
AG Theoretische Informatik

Übungsblatt 5

Ausgabe: 15.05.18
Abgabe: 22.05.18

Für ein Wort $w = w_1 \cdot w_2 \cdots w_m \in \Sigma^m$ sei $w^R := w_m \cdots w_2 \cdot w_1$.
Aufgabe 5.4 ist eine Bonusaufgabe mit 8 Extrapunkten.

Aufgabe 5.1 *Fundamentale Sprachklassen* (2+2+2+2+2 = 10 Punkte)

Für jede der folgenden Sprachen L ist die kleinste Sprachenklasse $\mathcal{K}$ anzugeben, zu der L gehört, bzw. zu zeigen, dass L nicht kontextfrei ist. Zur Auswahl stehen die Klassen REG der regulären Sprachen, DCFL der deterministisch kontextfreien Sprachen, ECFL der eindeutig kontextfreien Sprachen und CFL der kontextfreien Sprachen.

Ein informelles, aber natürlich überzeugendes Argument, warum $L \in \mathcal{K}$ gilt, ist ausreichend. Ein Nachweis, dass L nicht zu der nächstkleineren Sprachklasse gehört, ist nicht notwendig.

a) Sei $\Sigma = \{a, b\}$ und $L := \{\, a^n b^n \;:\; n \in \mathbb{N}\} \cup \{\, a^n b^{2n} \;:\; n \in \mathbb{N}\}$.

b) Sei $\Sigma = \{a, b, c, d\}$ und $L := \{\, a^i b^j c^i d^j \;:\; i, j \in \mathbb{N}\,\}$.

c) Sei $\Sigma = \{a, b, c, d\}$ und $L := \{\, a^i b^j c^j d^i \;:\; i, j \in \mathbb{N}\,\} \cup \{\, a^i b^i c^j d^j \;:\; i, j \in \mathbb{N}\,\}$.

d) Sei L die Komplementsprache von $\{\, u \in \{a, b\}^* \;:\; u = u^R\,\}$.

e) Sei $\Sigma = \{0, 1\}^3$. Für ein Wort $u \in \{0, 1\}^*$ sei Zahl(u) die natürliche Zahl mit Binärdarstellung u^R: das niedrigstwertige Bit steht also ganz links. Wir setzen

$$L := \big\{(u_0, v_0, w_0) \cdot (u_1, v_1, w_1) \cdots (u_n, v_n, w_n) \cdot (0, 0, w_{n+1}) : n \in \mathbb{N},\ \text{Zahl}(w) = \text{Zahl}(u) + \text{Zahl}(v)\big\}.$$

Aufgabe 5.2 *CYK und lineare Grammatiken* (6 Punkte)

Eine Grammatik $G = (\Sigma, V, S, P)$ heißt *linear*, wenn jede Produktion in P die Form $A \to cBd$, für $A, B \in V$ und $c, d \in \Sigma^*$ oder die Form $A \to a$ (für einen Buchstaben $a \in \Sigma$) hat.

Modifizieren Sie den Algorithmus von Cocke-Younger-Kasami (CYK), so dass das Wortproblem für eine gegebene lineare Grammatik $G = (\Sigma, V, S, P)$ in Zeit $\mathcal{O}(n^2)$ gelöst wird.

Analysieren Sie die asymptotische Laufzeit in Abhängigkeit von der Wortlänge $n := |w|$ einer Eingabe w. (Insbesondere können Sie die Anzahl $|P|$ der Produktionen, die Anzahl $|V|$ der Variablen wie auch die Längen der rechten Seiten von Produktionen $A \to cBd$ in P als Konstanten interpretieren.)

Hinweis: Der CYK-Algorithmus benutzt das Algorithmen-Paradigma der dynamischen Programmierung, um das „Masterproblem“ mit Hilfe einer Hierarchie von Teilproblemen zu lösen. Welche Teilprobleme löst der Algorithmus und wie viel Zeit reicht pro Teilproblem aus?

Aufgabe 5.3 *Was können DPDAs nicht?* (2+4+2 = 8 Punkte)

Wie zeigt man, dass eine Sprache L nicht deterministisch kontextfrei ist? Wenn bekannt ist, dass die Komplementsprache $\Sigma^* \setminus L$ nicht kontextfrei ist, dann kann L nicht deterministisch kontextfrei sein: Die Klasse der deterministisch kontextfreien Sprachen ist ja unter Komplementbildung abgeschlossen. Hier ist ein zweites Kriterium:

> Wenn jede Nerode-Klasse von L nur endlich viele Vertreter besitzt,
> dann ist L nicht deterministisch kontextfrei.

Man kann zeigen, dass alle Nerodeklassen der Sprache $L := \{w \in \{a,b\}^* \,:\, w = w^R\}$ ein-elementig sind. L ist also nicht deterministisch kontextfrei.

Sei $A = (Q, \Sigma, \Gamma, q_0, Z_0, \delta, F)$ ein DPDA und es gelte $|\Sigma| \geq 2$.

a) Für ein Wort $w \in \Sigma^*$ sei $s(w) \in \Sigma^*$ ein Wort, sodass der Keller von A nach Verarbeitung[1] von $w \cdot s(w)$ eine *minimale* Höhe unter allen Worten in $w \cdot \Sigma^*$ besitzt.

   Zeigen Sie: Für alle Suffixe z hängt die Berechnung von A auf der Eingabe $w \cdot s(w) \cdot z$ nur von dem Zustand ab, der nach Lesen von $w \cdot s(w)$ erreicht wird, und von dem zu diesem Zeitpunkt zuoberst liegenden Kellersymbol – nicht aber vom Inhalt des restlichen Kellers.

b) Zeigen Sie, dass es eine *unendliche* Teilmenge $U \subseteq \Sigma^*$ mit den folgenden Eigenschaften gibt.

   - Alle Worte $u \in U$ haben die Form $u = w \cdot s(w)$ (für ein Wort $w \in \Sigma^*$).
   - Es gibt einen Zustand $q \in Q$ und ein Kellersymbol $\gamma \in \Gamma$, sodass A *nach* Verarbeitung eines jeden Wortes $u \in U$ den Zustand q annimmt und γ das zuoberst liegende Kellersymbol ist.
   - Es gilt entweder $U \subseteq L(A)$ oder $U \subseteq \Sigma^* \setminus L(A)$.

   Hinweis: Hier sollten Sie $|\Sigma| \geq 2$ ausnutzen.

c) Zeigen Sie: Wenn jede Nerode-Klasse von L nur endlich viele Vertreter besitzt, dann ist L nicht deterministisch kontextfrei.

Bitte wenden!

[1] Wir sagen, dass A eine Eingabe verarbeitet hat, wenn A die Eingabe komplett gelesen hat und einen neuen Buchstaben lesen möchte.

Aufgabe 5.4 ***Inhärente Mehrdeutigkeit*** (Bonusaufgabe: 2+3+3 = 8 Extrapunkte)

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 Ogdens 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 Ogdens 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 $S \stackrel{*}{\Rightarrow} u_1 A y_1$ sowie $A \stackrel{*}{\Rightarrow} v_1 A x_1 \mid w_1$.

Wir schauen noch etwas genauer hin.

i) Zeigen Sie: $v_1 = a^j$ und $x_1 = b^j$ für ein j mit $1 \le j \le M$. 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 $S \stackrel{*}{\Rightarrow} u_1 A y_1$ und $A \stackrel{*}{\Rightarrow} v_1 A x_1 \mid w_1$ abgeleitet werden. Es werden mehr als $M!$ viele b's durch die Ableitung $A \stackrel{*}{\Rightarrow} v_1 A x_1$ 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 $1 \le k \le M$) sowie eine Variable B mit den Ableitungen $S \stackrel{*}{\Rightarrow} u_2 B y_2$ und $B \stackrel{*}{\Rightarrow} v_2 B x_2 \mid w_2$. 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.