Institut für Informatik
M. Mutz

Übungsblatt 6

Theoretische Informatik

Bearbeitungszeit:	20.05.2024 bis 26.05.2024, 16:00 Uhr
Besprechung:	27.05.2024, 10:30 Uhr in Hörsaal 5E

Abgabe: als PDF über das ILIAS
Gruppenabgaben möglich und erwünscht

Aufgabe 1 (CYK-Algorithmus I) 10 Punkte
Gegeben sei die Grammatik $G' = (\Sigma, N', S, P')$ mit $N' = \{S, A, B, C, D_1, D_2, X_0, X_1\}$ und

$$\begin{aligned} P' = \{S &\rightarrow X_0C \mid X_0X_0, \\ C &\rightarrow X_0A, \\ A &\rightarrow X_1D_1 \mid X_1D_2 \mid X_0B \mid 0, \\ D_1 &\rightarrow AD_2, \\ D_2 &\rightarrow X_1X_1, \\ B &\rightarrow X_0B \mid 0, \\ X_0 &\rightarrow 0, \\ X_1 &\rightarrow 1\}. \end{aligned}$$

a) Entscheiden Sie mit dem Algorithmus von Cocke, Younger und Kasami (CYK-Algorithmus), ob die Wörter

   i) $w_1 = 0011$ und
   ii) $w_2 = 001011$

   in $L(G')$ liegen. Geben Sie die Tabelle bzw. Dreiecksmatrix dabei jeweils vollständig an. *Bitte beachten Sie, dass Ihre Grammatik in CNF sein muss, damit Sie den CYK-Algorithmus anwenden können.*

b) Geben Sie $L(G')$ formal als Menge von Wörtern an. Die Angabe soll als in sich abgeschlossener Mengenausdruck geschehen. Referenzieren Sie keine weiteren Konstrukte wie die Grammatik G', Teile von G', wie die Produktionsregeln P', einen regulären Ausdruck, einen Automaten, oder eine andere Grammatik. Die Ausnahme hier stellt Σ dar, welches Sie referenzieren dürfen. Lösungen wie $L(G') = \{w|w \in L(...)\}$ führen zu 0 Punkten.

Lösungsvorschlag:

a) (i) :

i	1	2	3	4
3	D_1			
2		D_1		
1	S, C, A, B		D_2	
0	X_0, B, A	X_0, B, A	X_1	X_1
j	0	0	1	1

$0011 \notin L(G)$, da D_1 nicht das Startsymbol ist.

(ii) :

i	1	2	3	4	5	6
5	S					
4		C				
3			A			
2				D_1		
1	S, A, B, C				D_2	
0	X_0, B, A	X_0, B, A	X_1	X_0, B, A	X_1	X_1
j	0	0	1	0	1	1

$001011 \in L(G)$, da sich S im linken oberen Kästchen befindet und S das Startsymbol ist.

b) $L(G) = \{0^2 1^n 0^m 1^{2n} \mid m, n \geq 0\}$

Aufgabe 2 (CYK-Algorithmus II) 10 Punkte

Gegeben seien das Alphabet $\Sigma = \{a, b, c\}$ und die folgende Tabelle, welche mit dem CYK-Algorithmus für das Wort $w = aabbcc$ erstellt wurde. Aus dem Ergebnis kann gefolgert werden, dass w in der Sprache liegt, welche die zugrundeliegende Grammatik erzeugt.

i	1	2	3	4	5	6
5	S, Z					
4	Y	Y				
3	X		Z, S			
2	A		Y	Y		
1		X		Z	Z, S	
0	A	A	B	B	Y	Y
j	a	a	b	b	c	c

Geben Sie eine mögliche Grammatik G an, welche diese Tabelle erzeugen kann.

Lösungsvorschlag:
Eine mögliche Grammatik ist
$G = (\Sigma, N, S, P)$ mit $\Sigma = \{a, b, c\}, N = \{S, A, B, X, Y, Z\}, S, P)$ und

$$\begin{aligned}
P = \{S &\to YY, \\
Y &\to BZ \mid AZ, \\
A &\to AX \mid BX, \\
Z &\to BY \mid YY, \\
X &\to AB, \\
Y &\to c, \\
B &\to b, \\
A &\to a\}.
\end{aligned}$$

Hinweis: Hier sind weitere Grammatiken denkbar. Auch Z kann z.B. das Startsymbol sein. Außerdem sind weitere Regeln möglich, die rechte Seiten haben, die nicht in der oben angegebenen Grammatik vorkommen.

Aufgabe 3 (Kellerautomat I)10P
Der Kellerautomat $M = (\{a, b\}, \{\#, A, B\}, \{z_0, z_1, z_2, z_3, z_4\}, \delta, z_0, \#)$ verfügt über die folgenden Regeln um δ zu definieren:

$$\begin{array}{rrr}
z_0a\# \to z_1AA\# & z_1bA \to z_1\lambda & z_2\lambda\# \to z_0\# \\
z_0b\# \to z_2B\# & z_1\lambda\# \to z_0\# & z_3\lambda B \to z_2\lambda \\
z_0\lambda\# \to z_4\lambda & z_2aB \to z_3\lambda & z_3\lambda\# \to z_1A\# \\
z_1aA \to z_1AAA & z_2bB \to z_2BB &
\end{array}$$

a) Liegt das Wort *bababb* in der Sprache des Kellerautomaten $L(M)$? Verwenden Sie in Ihrer Argumentation die entsprechende(n) Konfigurationsfolge(n).

b) Geben Sie jeweils die Inhalte des Kellers an, nachdem M die folgenden Zeichenketten gelesen hat:

i) $b^3ab^4a^3$

ii) a^2b^3

c) Beschreiben Sie die Sprache $L(M)$ möglichst formal. Die Angabe soll als in sich abgeschlossener Mengenausdruck geschehen. Referenzieren Sie keine weiteren Konstrukte wie den Kellerautomaten M, Teile von M, wie die Überführungsfunktion δ, die Relation von Konfigurationenfolgen, einen regulären Ausdruck, eine Grammatik, oder einen anderen Automaten. Die Ausnahme hier stellt $\Sigma = \{a, b\}$ dar, welches Sie referenzieren dürfen. Lösungen wie $L(M) = \{w | w \in L(...)\}$ führen zu 0 Punkten.

d) Zeigen Sie, dass die Überführungsfunktion δ von M nicht deterministisch ist.

Lösungsvorschlag:

a) Eine mögliche Konfigurationenfolge für das Wort *bababb*, die auch zeigt, dass das Wort in der Sprache L(M) enthalten ist, ist folgende:

$$\begin{aligned}(z_0, bababb, \#) &\vdash_M (z_2, ababb, B\#) \vdash_M (z_3, babb, \#) \vdash_M (z_1, babb, A\#)\\ &\vdash_M (z_1, abb, \#) \vdash_M (z_0, abb, \#) \vdash_M (z_1, bb, AA\#)\\ &\vdash_M (z_1, b, A\#) \vdash_M (z_1, \lambda, \#) \vdash_M (z_0, \lambda, \#) \vdash_M (z_4, \lambda, \lambda)\end{aligned}$$

b) Wir führen die Konfigurationenfolgen für die Wörter $b^3ab^4a^3$ und a^2b^3 bis wir jeweils das Wort abgearbeitet haben und dann geben wir den jeweiligen Inhalt des Kellers an.
Die Konfigurationenfolge für das Wort $b^3ab^4a^3$:

$$\begin{aligned}(z_0, b^3ab^4a^3, \#) &\vdash_M (z_2, b^2ab^4a^3, B\#) \vdash_M (z_2, bab^4a^3, BB\#) \vdash_M (z_2, ab^4a^3, BBB\#)\\ &\vdash_M (z_3, b^4a^3, BB\#) \vdash_M (z_2, b^4a^3, B\#) \vdash_M (z_2, b^3a^3, BB\#)\\ &\vdash_M (z_2, b^2a^3, BBB\#) \vdash_M (z_2, ba^3, BBBB\#) \vdash_M (z_2, a^3, BBBBB\#)\\ &\vdash_M (z_3, a^2, BBBB\#) \vdash_M (z_2, a^2, BBB\#) \vdash_M (z_3, a, BB\#)\\ &\vdash_M (z_2, a, B\#) \vdash_M (z_3, \lambda, \#)\\ (&\vdash_M (z_1, \lambda, A\#)\ )\end{aligned}$$

Der Inhalt des Kellers nach dem Lesen der Zeichenkette $b^3ab^4a^3$ ist also $\#$.

Hier ist es möglich mit dem λ-Übergang noch einen Schritt mehr zu machen. Mit und ohne diesem zusätzlichen Schritt wird die Aufgabenstellung erfüllt.

Die Konfigurationenfolge für das Wort a^2b^3:

$$\begin{aligned}(z_0, a^2b^3, \#) \quad &\vdash_M \quad (z_1, ab^3, AA\#) \vdash_M (z_1, b^3s, AAAA\#) \vdash_M (z_1, b^2, AAA\#) \\ &\vdash_M \quad (z_1, b, AA\#) \vdash_M (z_1, \lambda, A\#)\end{aligned}$$

Der Inhalt des Kellers nach dem Lesen der Zeichenkette a^2b^3 ist also $A\#$.

c) $L(M) = \{w \in \{a,b\}^* \mid 2 \cdot |w|_a = |w|_b\}$, wobei $|w|_a$ für die Anzahl der a's im Wort w steht und $|w|_b$ für die Anzahl der b's im Wort w.

d) M ist kein deterministischer Kellerautomat, weil die Ungleichungen
$||\delta(z_0, a, \#)|| + ||\delta(z_0, \lambda, \#)|| = 1 + 1 = 2 > 1$
und
$||\delta(z_0, b, \#)|| + ||\delta(z_0, \lambda, \#)|| = 1 + 1 = 2 > 1$
gelten.
(hier muss natürlich nur eines davon angegeben werden)

Aufgabe 4 (Abschlusseigenschaften)10P

(a) Betrachten Sie eine Sprache L und einen DFA $M = (\Sigma, Z, \delta, z_0, F)$ mit totaler Überführungsfunktion δ und mit $L = L(M)$.
Geben Sie einen DFA M' mit $\overline{L} = L(M')$ als Tupel an, um zu zeigen, dass die Klasse der regulären Sprachen unter Komplement abgeschlossen ist.
Hinweis: Ihr DFA M' soll nicht konkret angegeben werden, sondern muss aus jedem beliebigen Automaten M, wie er in der Aufgabenstellung angegeben ist, konstruierbar sein.

(b) Zeigen Sie nur mit Hilfe von Abschlusseigenschaften für kontextfreie Sprachen und Satz 3.36 folgende Behauptungen. Sie dürfen Sprachen, die von regulären Ausdrücken erzeugt werden, direkt als regulär verwenden. Aus der Vorlesung bekannte Sprachen dürfen wieder direkt verwendet werden (z.B. bewiesene Behauptungen aus dem Skript zu spezifischen Sprachen sind also auch erlaubt zu nutzen).

$$L_1 = \{a^i b^{i+j} a^j \mid i, j \geq 0\} \in CF \text{ mit } L_1 \subseteq \{a,b\}^*$$
$$L_2 = \{a^n b^m c^{max(n,m)} \mid n, m \geq 0\} \notin REG \text{ mit } L_2 \subseteq \{a,b,c\}^*$$

Lösungsvorschlag:

(a) Der Automat muss genau die Wörter akzeptieren, die M nicht akzeptiert. Daher sind alle Zustände Endzustände, die nicht Endzustände in M waren. Somit ist $M' = (\Sigma, Z, \delta, z_0, Z \setminus F)$.

Falls δ als nicht total angenommen wird, muss ein Zustand z' mit $z' \notin Z$ hinzugenommen werden und δ wird so zu δ' ergänzt, dass $\delta'(z', x) := z'$ für alle $x \in \Sigma$ und $\delta'(z_i, x) := z_j$ für alle $z_i, z_j \in Z, x \in \Sigma$ mit $\delta(z_i, x) = z_j$, sowie $\delta'(z_i, x) := z'$ für alle $z_i \in Z, x \in \Sigma$ mit $\delta(z_i, x)$ nicht definiert.
Damit ist dann $M' = (\Sigma, Z', \delta', z_0, Z' \setminus F)$ mit $Z' = Z \cup \{z'\}$

(b) L_1: Es ist bekannt: $L_1' = \{a^i b^i | \; i \geq 0\} \in CF$. Dann ist $L_1 = L_1' sp(L_1')$ (Konkatenation und Spiegelung) und damit kontextfrei.

L_2: Wir wissen, dass $\{a^n b^n \mid n \geq 1\} \in CF$ ist. $\{c^m \mid m \geq 1\}$ ist ebenfalls kontextfrei. Mit den Abschlusseigenschaften der Kontextfreien Sprachen ist somit auch die Konkatenation $\{a^n b^n c^m \mid n, m \geq 1\} \in CF$.
Angenommen, L_2 wäre regulär. Dann gilt mit Satz 3.36 $\{a^n b^n c^m \mid n, m \geq 1\} \cap L_2 \in CF$. Aber $\{a^n b^n c^m \mid n, m \geq 1\} \cap L_2 = \{a^n b^n c^n \mid n \geq 1\} \notin CF$ nach Behauptung 3.20. Das ist ein Widerspruch, somit ist unsere Annahme falsch und es gilt $L_2 \notin REG$.