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

Übungsblatt 7

Vorlesung Theoretische Grundlagen der Informatik im WS 16/17

Ausgabe 17. Januar 2017
Abgabe 31. Januar 2017, 11:00 Uhr (im Kasten im UG von Gebäude 50.34)

Bitte nutzen Sie den WebInScribe Deckblattgenerator
und heften Sie das Deckblatt an Ihr Übungsblatt.
`https://webinscribe.ira.uka.de/deckblatt/index.php?course=10588`.

Aufgabe 1 (2 + 2 = 4 Punkte)

Zeigen Sie, dass folgende Sprachen nicht kontextfrei sind:

(a) $L_1 = \{w \in \{a, b, c\}^\star : |w|_a = |w|_b = |w|_c\}$, wobei $|w|_x$ angibt, wie oft x in w vorkommt.

(b) $L_2 = \left\{a^i b^i c^j \mid i \geq j\right\}$.

Lösung:

(a) Idee: widerlege die Annahme, dass L_1 kontextfrei ist, mit dem Pumpinglemma. Nehme also an, dass L_1 kontextfrei ist. Dann existiert nach dem Pumpinglemma ein $n \in \mathbb{N}$, so dass für jedes Wort z mit $|z| \geq n$ eine Zerlegung $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$, so dass $uv^i wx^i y \in L_1$ für alle $i \geq 0$. Wegen $|vwx| \leq n$ kann vx kann nicht gleichzeitig a's und c's enthalten. Das Wort uv^2wx^2y enthält also nicht gleich viele a's und c's, liegt also nicht in L_1, was ein Widerspruch zur Annahme ist, dass L_1 kontextfrei ist.

(b) Idee: widerlege die Annahme, dass L_2 kontextfrei ist, mit Ogdens Lemma. Wähle n als Konstante aus Ogdens Lemma. Betrachte $z = a^{n+1}b^{n+1}c^{n+1} \in L_2$ und markiere alle b's. Betrachte Zerlegung $z = uvwxy$ gemäß Ogdens Lemma. Dann muss mindestens ein b zu vx gehören. Da höchstens n markierte Buchstaben zu vwx gehören dürfen, kann vx nicht gleichzeitig a's und c's enthalten. Das Wort uv^0wx^0y enthält dann entweder zu viele c's, oder zu viele b's und ist deswegen nicht in L_2 enthalten. Widerspruch.

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

- Über dem Alphabet $\Sigma = \{a, b\}$ sei der reguläre Ausdruck $r := (a \cup (ab(b)^*ba))^*$ gegeben. Gebe eine Typ-k Grammatik mit maximalem k an, die $L(r)$ erzeugt.

  Lösung:
  Die Sprache $L(r)$ ist regulär und kann somit von einer Typ-3 Grammatik erzeugt werden. $\Sigma = \{a, b\}$, Variablen $\{S, B_1, B_2, B_3, A, T\}$, Startsymbol S

$$\begin{aligned} S &\rightarrow aS|aB_1|\epsilon \\ B_1 &\rightarrow bB_2|bB_3 \\ B_2 &\rightarrow bB_2|bB_3 \\ B_3 &\rightarrow bA \\ A &\rightarrow aT|aS \\ T &\rightarrow \epsilon \end{aligned}$$

- Konstruiere eine Typ-1-Grammatik, welche die Sprache

$$L = \{a^n b^{2n} c^n \ : \ n \geq 1\}$$

  erzeugt.

  Lösung:

  $\Sigma = \{a, b, c\}$, Variablen $\{S, A, T, C, U, V\}$, Startsymbol S

$$\begin{aligned} S &\rightarrow ATC && (1) \\ T &\rightarrow ATC && (2) \\ T &\rightarrow UV && (3) \\ AU &\rightarrow aUb && (4) \\ Aa &\rightarrow aA && (5) \\ VC &\rightarrow bVc && (6) \\ cC &\rightarrow Cc && (7) \\ U &\rightarrow \epsilon && (8) \\ V &\rightarrow \epsilon && (9) \end{aligned}$$

- Konstruiere eine Typ-2-Grammatik, welche die Sprache

$$L = \{w \in \{0, 1\}^* \ : \ |w|_0 = |w|_1\}$$

  erzeugt. Zeigen Sie, dass Ihre Grammatik tatsächlich die Sprache L erzeugt.

  Lösung:
  Die folgende Grammatik $G = (\Sigma, V, S, R)$ mit

  - Alphabet $\Sigma = \{0, 1\}$,
  - Variablen $V = \{S\}$,
  - Startsymbol S,
  - Produktionen $R = \{S \rightarrow 0S1|1S0|SS|\epsilon\}$

  ist vom Typ 2.

  Noch zu Zeigen: $L(G) = L$

- $L(G) \subseteq L$. Gilt weil jede Regel entweder kein Element aus Σ oder genau eine 0 und genau eine 1 erzeugt.

- $L(G) \supseteq L$. Sei w ein Wort aus L. Wir beweisen durch strukturelle Induktion, dass w aus S durch Produktionen aus R erzeugt werden kann.

- **Induktionshypothese:** Alle Wörter aus L mit Länge maximal n sind aus S durch die Produktionen R erzeugbar.
- **Induktionsschluss:** Alle Wörter aus L mit Länge maximal $n+2$ sind aus S durch die Produktionen R erzeugbar.

Sei $w \in L$ mit Länge $n+2$

- 1. Fall $w = 0\tilde{w}1$ bzw $w = 1\tilde{w}0$ für ein w. Dann wende die Regel $S \to 0S1$ bzw $S \to 1S0$ an. Wegen der Induktionshypothese ist $\tilde{w}$ erzeugbar und die Behauptung gilt
- 2. Fall $w = 0\tilde{w}0$. Wegen $|w|_0 = |w|_1$ gilt $|\tilde{w}|_0 + 2 = |\tilde{w}|_1$. Deswegen gibt es eine Zerlegung $w = uv$, $u \neq \epsilon, v \neq \epsilon$ mit $|u|_0 = |u|_1$ und $|v|_0 = |v|_1$. Finde eine solche Zerlegung (geht in linearer Zeit), wende die Regel $S \to SS$ and. Mit der Induktionshypothese folgt die Behauptung.
- 3. Fall $w = 1\tilde{w}1$. Geht analog zum 2. Fall.

Aufgabe 3 (2+2 = 4 Punkte)

Sind die Sprachen von Typ 0 abgeschlossen unter

(a) Vereinigung, Durchschnitt, Komplementbildung,

(b) Konkatenation?

Lösung:

(a) Aus VL bekannt: Chomsky-Typ-0 entspricht genau den semientscheidbaren Sprachen. Aus Übungsblatt bekannt: Semi-entscheidbare Sprachen sind abgeschlossen unter $\cup$, $\cap$ aber nicht abgeschossen unter Komplement

(b) Chomsky-Typ-0 Sprachen sind abgeschlossen unter Konkatenation. Seien $G_1 = (\Sigma_1, V_1, S_1, P_1)$ und $G_2 = (\Sigma_2, V_2, S_2, P_2)$ Typ-0 Grammatiken. O.B.d.A sein $\Sigma_1, V_1, \Sigma_2, V_2$ disjunkt. Die Grammatik $G = (\Sigma_1 \cup \Sigma_2, V_1 \cup V_2, S, P_1 \cup P_2 \cup \{S \to S_1 S_2\})$ erzeugt $G_1 \cdot G_2$.

Falls $\Sigma_1, V_1, \Sigma_2, V_2$ nicht disjunkt sind, benenne die Mengen konsistent um, um dies sicher zu stellen.

Aufgabe 4 (4 Punkte)

Sei $G = (\Sigma, V, S, R)$ die CH-2-Grammatik mit $\Sigma = \{a, b\}$, $V = \{A, B, C, D, E, S\}$ und der folgenden Regelmenge R:

$$\begin{array}{rcl} S & \to & A \mid aAa \mid bBb \mid \varepsilon \\ A & \to & C \mid a \\ B & \to & C \mid b \\ C & \to & CDE \mid \varepsilon \\ D & \to & A \mid B \mid ab \\ E & \to & S \end{array}$$

Bestimme eine Grammatik G' für $L(G)$ in Chomsky-Normalform.

Lösung:
1. Schritt: Alle Regeln enthalten auf der rechten Seite nur Symbole aus V oder nur ein Symbol aus Σ.

$$\begin{array}{rcl} S & \to & A \mid Y_aAY_a \mid Y_bBY_b \mid \varepsilon \\ A & \to & C \mid a \\ B & \to & C \mid b \\ C & \to & CDE \mid \varepsilon \\ D & \to & A \mid B \mid Y_aY_b \\ E & \to & S \\ Y_a & \to & Y_a \\ Y_b & \to & Y_b \end{array}$$

2. Schritt: Alle rechten Seiten haben Länge ≤ 2.

$$\begin{array}{rcl} S & \to & A \mid Y_aC_1 \mid Y_bC_2 \mid \varepsilon \\ A & \to & C \mid a \\ B & \to & C \mid b \\ C & \to & CC_3 \mid \varepsilon \\ D & \to & A \mid B \mid Y_aY_b \\ E & \to & S \\ C_1 & \to & AY_a \\ C_2 & \to & BY_b \\ C_3 & \to & DE \\ Y_a & \to & a \\ Y_b & \to & b \end{array}$$

3. Schritt: Es kommen keine Regeln $A \to \varepsilon$ vor. Berechne $V' = \{A \in V \mid A \overset{*}{\to} \varepsilon\} =$

$\{S, C, A, B, D, E\}$.

$$\begin{array}{rcl}
S & \to & A \mid Y_aC_1 \mid Y_bC_2 \\
A & \to & C \mid a \\
B & \to & C \mid b \\
C & \to & CC_3 \mid C_3 \\
D & \to & A \mid B \mid Y_aY_b \\
E & \to & S \\
C_1 & \to & AY_a \mid Y_a \\
C_2 & \to & BY_b \mid Y_b \\
C_3 & \to & DE \mid E \\
Y_a & \to & a \\
Y_b & \to & b
\end{array}$$

4. Schritt: Ersetzung aller Kettenregeln $A \to B$. Der Abhängigkeitsgraph aller an Kettenregeln beteiligten Variablen ist gegeben durch: ...

Es gibt nur einen Kreis $S \to A \to C \to C_3 \to E$ mit beteiligten Kettenregeln $S \to A, A \to C, C \to C_3, C_3 \to E$. Diese Regeln werden gelöscht und alle Vorkommen von S, A, C, C_3, E in allen Regeln werden durch S ersetzt. Wir erhalten die Regelmenge:

$$\begin{array}{rcl}
S & \to & S \mid Y_aC_1 \mid Y_bC_2 \\
S & \to & S \mid a \\
B & \to & S \mid b \\
S & \to & SS \mid S \\
D & \to & S \mid S \mid Y_aY_b \\
S & \to & S \\
C_1 & \to & SY_a \mid Y_a \\
C_2 & \to & BY_b \mid Y_b \\
S & \to & DS \mid S \\
Y_a & \to & a \\
Y_b & \to & b
\end{array}$$

Nach Löschen aller Regeln der Form $S \to S$ erhalten wir dann:

$$\begin{array}{rcl}
S & \to & Y_aC_1 \mid Y_bC_2 \mid a \mid SS \mid DS \\
B & \to & S \mid b \\
D & \to & S \mid Y_aY_b \\
C_1 & \to & SY_a \mid Y_a \\
C_2 & \to & BY_b \mid Y_b \\
Y_a & \to & a \\
Y_b & \to & b
\end{array}$$

Der Abhängigkeitsgraph aller an Kettenregeln beteiligten Variablen ist nun azyklisch und kann topologisch geordnet werden. Eine mögliche Ordnung ist gegeben durch $B, D, S, C_1, Y_a, C_2, Y_b$. Wir verarbeiten die Regeln in umgekehrter Reihenfolge und ersetzen Kettenregeln.

$$
\begin{aligned}
S &\rightarrow Y_aC_1 \mid Y_bC_2 \mid a \mid SS \mid DS \\
B &\rightarrow Y_aC_1 \mid Y_bC_2 \mid a \mid SS \mid DS \mid b \\
D &\rightarrow Y_aC_1 \mid Y_bC_2 \mid a \mid SS \mid DS \mid Y_aY_b \\
C_1 &\rightarrow SY_a \mid a \\
C_2 &\rightarrow BY_b \mid b \\
Y_a &\rightarrow a \\
Y_b &\rightarrow b
\end{aligned}
$$

Diese Regelmenge ist in Chomsky-Normalform. Zum Schluß müssen wir die Grammatik noch ergänzen durch die Regeln $S' \rightarrow \varepsilon$ und $S' \rightarrow S$ für ein neues Startsymbol S', da die ursprüngliche Grammatik die Ableitung $S \overset{*}{\rightarrow} \varepsilon$ zuläßt.

Aufgabe 5

(1 + 1 + 0,5 + 1,5 + 1 = 5 Punkte)

(a) Bestimme eine möglichst kurze Grammatik G_R für reguläre Ausdrücke über dem Alphabet $\Sigma = \{0, 1\}$ mit maximalem Chomsky-Typ.

(b) Bestimme einen Ableitungsbaum für das Wort $(1 \cup 0)^\star 1$.

(c) Ist die Grammatik eindeutig?

(d) Bringen Sie G_R in Chomsky-Normalform.

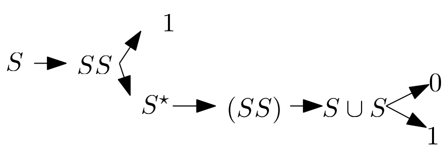
(e) Überprüfen Sie die Zugehörigkeit des Worts $(1 \cup 0)^\star 1$ zu $L(G_R)$ mit dem CYK-Algorithmus.

Lösung:

(a) Grammatik G_R für reguläre Ausdrücke:

$$S \rightarrow S^\star \mid S \cup S \mid SS \mid (S) \mid \varepsilon \mid 0 \mid 1$$

(b) Ableitungsbaum



(c) Es gibt mehrere Wege das Wort 1 (bilde SS auf $\varepsilon 1$ ab oder umgekehrt) aus der Grammatik herzuleiten, daher ist die Grammatik inhärent mehrdeutig.

(d) Normalisierung: 1. Schritt: Alle Regeln enthalten auf der rechten Seite nur Symbole aus V oder nur ein Symbol aus Σ.

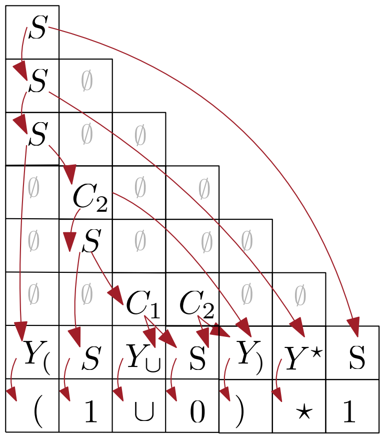
$$
\begin{aligned}
S &\rightarrow SY_* \mid SY_\cup S \mid SS \mid Y_(SY_) \mid \varepsilon \mid 0 \mid 1 \\
Y_* &\rightarrow * \\
Y_\cup &\rightarrow \cup \\
Y_(&\rightarrow (\\
Y_) &\rightarrow)
\end{aligned}
$$

2. Schritt: Alle rechten Seiten haben Länge ≤ 2.

$$
\begin{array}{rcl}
S & \to & SY_* \mid SC_1 \mid SS \mid Y_(C_2 \mid \varepsilon \mid 0 \mid 1 \\
C_1 & \to & Y_\cup S \\
C_2 & \to & SY_) \\
Y_* & \to & * \\
Y_\cup & \to & \cup \\
Y_(& \to & (\\
Y_) & \to &)
\end{array}
$$

Grammatik ist schon in Chomsky-Normalform.

(e) CYK-Tabelle:



Aufgabe 6

(1 + 4 = 5 Punkte)

Gegeben sei die Grammatik $G_1 = (\Sigma, V, S, R)$ mit $\Sigma = \{a, b\}$, $V = \{S, A, B\}$ und der Regelmenge R:

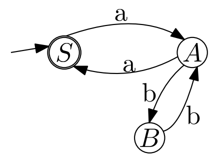
$$
\begin{array}{rcl}
S & \to & aA \mid \varepsilon \\
A & \to & aS \mid bB \\
B & \to & bB \mid bA\,.
\end{array}
$$

(a) Welchen Chomsky-Typ hat G_1?

(b) Konstruiere einen nichtdeterministischen endlichen Automaten für $L(G)$ mithilfe des Verfahrens aus dem Skript.

Lösung:

(a) G_1 ist eine CH-2-Grammatik.



(b)