FGI-1 – Formale Grundlagen der Informatik I

Logik, Automaten und Formale Sprachen

Musterlösung 4 : Kontextfreie Sprachen und Grammatiken

Präsenzteil am 28.-30. April – Abgabe am 5.-8. Mai 2015

Präsenzaufgabe 4.1: Gegeben sei die Grammatik G mit den Produktionen

$$\begin{array}{rcl} S & \longrightarrow & AS \mid V \\ X & \longrightarrow & uX \mid S \\ Y & \longrightarrow & uAY \mid YX \\ A & \longrightarrow & XY \mid SV \\ V & \longrightarrow & v \end{array}$$

Konstruieren Sie zu G eine äquivalente, reduzierte Grammatik G' nach dem Verfahren aus der Vorlesung (siehe auch Theorem 9.11 im Skript).

Lösung: Wir verfahren wie im Beweis von Theorem 9.11. Zunächst ermitteln wir die produktiven Symbole durch

$$\begin{array}{rcl} M_0 & = & \{u, v\} \\ M_1 & = & \{u, v, V\} \\ M_2 & = & \{u, v, V, S\} \\ M_3 & = & \{u, v, V, S, X, A\} \end{array}$$

Wegen $M_3 = M_4$ bricht das Verfahren ab. Durch Einschränkung der Produktionen auf $P \cap (M_3 \times M_3^*)$ erhalten wir nun die Grammatik G_2 mit den Produktionen

$$\begin{array}{rcl} S & \longrightarrow & AS \mid V \\ X & \longrightarrow & uX \mid S \\ A & \longrightarrow & SV \\ V & \longrightarrow & v \end{array}$$

Im zweiten Schritt ermitteln wir die erreichbaren Nonterminale zu

$$\begin{array}{rcl} M_0 & = & \{S\} \\ M_1 & = & \{S, A, V\} \end{array}$$

Wegen $M_1 = M_2$ bricht das Verfahren ab. Wir erhalten nun die Grammatik G' mit der Produktionenmenge

$$\begin{array}{rcl} S & \longrightarrow & AS \mid V \\ A & \longrightarrow & SV \\ V & \longrightarrow & v \end{array}$$

Nach Theorem 9.11 gilt $L(G') = L(G)$.

Man kann G' noch weiter vereinfachen, indem man die rechten Seiten der Produktionen von A und V direkt einsetzt. Man erhält so:

$$S \longrightarrow SvS \mid v$$

und kann nun sehen, dass die generierte Sprache gerade die Menge der Worte aus v^+ ist, die eine ungerade Länge haben.

Präsenzaufgabe 4.2: Nutzen Sie das Pumping Lemma für kontextfreie Sprachen, um zu zeigen, dass

$$L := \{a^i b^j c^k \mid i \leq j \leq k\}$$

nicht kontextfrei ist.

Lösung: Angenommen L wäre kontextfrei. Sei n die Zahl aus dem Pumping Lemma. Wir betrachten das Wort $z = a^n b^n c^n$. Offensichtlich ist $z \in L$ und $|z| \geq n$. Sei nun $z = uvwxy$ eine Zerlegung von z, die die Bedingungen i und ii des Pumping Lemmas erfüllt, für die also $|vwx| \leq n$ und $|vx| \geq 1$ gilt. Wir führen nun jede dieser Zerlegungen zu einem Widerspruch mit der dritten Eigenschaft des Pumping Lemmas.

Zunächst gibt es wegen $|vwx| \leq n$ und der Form von z fünf Fälle zu unterscheiden: $vwx \in \{a\}^*$, $vwx \in \{a\}^*\{b\}^*$, $vwx \in \{b\}^*$, $vwx \in \{b\}^*\{c\}^*$ und $vwx \in \{c\}^*$. In den Fällen $vwx \in \{a\}^*$ und $vwx \in \{b\}^*$ führt wegen $|vx| \geq 1$ die Betrachtung von uv^2wx^2y bereits zum Widerspruch. Es ist dann $uv^2wx^2y = a^{n+j}b^n c^n$ bzw. $uv^2wx^2y = a^n b^{n+j} c^n$ mit $j > 0$ und folglich gilt nicht mehr, dass der Exponent von c am größten ist. Ähnlich führt im Fall $vwx \in \{c\}^*$ die Betrachtung von $uv^0wx^0y = uwy = a^n b^n c^{n-j}$ (mit $j > 0$) zum Widerspruch.

Gilt in $vwx \in \{a\}^*\{b\}^*$ bzw. in $vwx \in \{b\}^*\{c\}^*$, dass v oder x bereits aus mehr als einem Symbol aufgebaut ist (also selbst in $\{a\}^+\{b\}^+$ bzw. $\{b\}^+\{c\}^+$ ist), so führt uv^2wx^2y sofort zu einem Widerspruch, da dieses Wort dann nicht einmal mehr in $\{a\}^*\{b\}^*\{c\}^*$ ist (und damit ganz bestimmt nicht in L).

Ist jedoch im Fall $vwx \in \{a\}^*\{b\}^*$ das v in den as und das x in den bs, so führt ähnlich wie bei den zuerst diskutierten Fällen die Betrachtung von $uv^2wx^2y = a^{n+j}b^{n+k}c^n$ zu einem Widerspruch ($n + j \leq n + k \leq n$ gilt nicht, da wegen $|vx| \geq 1$ auch $j + k \geq 1$ sein muss). Und ist im Fall $vwx \in \{b\}^*\{c\}^*$ das v in den bs und das x in den cs, so führt ebenfalls ähnlich zu den zuerst diskutierten Fällen die Betrachtung von $uv^0wx^0y = a^n b^{n-j} c^{n-k}$ zu einem Widerspruch (abermals gilt $n \leq n - j \leq n - k$ nicht, da $j + k \geq 1$ sein muss).

Damit sind alle Fälle zum Widerspruch geführt und es gibt also keine Zerlegung von z in $uvwxy$ derart, dass die drei Bedingungen des Pumping Lemmas erfüllt sind. Folglich ist die ursprüngliche Annahme, L wäre kontextfrei, nicht haltbar und L ist also nicht kontextfrei.

Übungsaufgabe 4.3: Sei

von
4

$$M := \{a^n b^m \mid n > m\}$$

1. Zeigen Sie mittels des Pumping Lemmas, dass M nicht regulär ist.
2. Konstruieren Sie eine kontextfreie Grammatik G mit $L(G) = M$ und zeigen Sie insbesondere, dass dies tatsächlich gilt.

Lösung:

1. Wir halten den Beweis diesmal knapp.

   Angenommen M wäre regulär. Dann gilt das Pumping Lemma. Sei nun k die Zahl aus dem Pumping Lemma. Wir betrachten $z := a^{k+1}b^k$. Es ist $|z| \geq k$ und $z \in M$. Sei nun $z = uvw$ eine Zerlegung, die $|uv| \leq k$ und $|v| \geq 1$ erfüllt (die ersten zwei Bedingungen des Pumping Lemmas). Es gilt dann wie oben $v = a^j$ für ein j mit $1 \leq j \leq k$. Wir betrachten nun das Wort $z' := uv^0w = a^{k+1-j}b^k$. Wegen $j \geq 1$ gilt $k + 1 - j \leq k$, woraus $z' \notin M$ folgt. Im Widerspruch zur dritten Eigenschaft des Pumping Lemmas, die ja $uv^iw \in M$ fordert, wobei auch $i = 0$ sein kann.

2. Wir geben die Produktionen an:

$$\begin{array}{rcl} S & \rightarrow & aSb \mid A \\ A & \rightarrow & aA \mid a \end{array}$$

   Das Startsymbol ist S, die Menge der Nonterminale $V_N = \{S, A\}$ und die Menge der Terminale $V_T = \{a, b\}$.

   Wir wollen nun noch zeigen, dass tatsächlich $L(G) = M$ gilt. Sei zunächst $w \in M$. Dann hat w die Form $a^n b^m$ mit $n > m$. In der Grammatik wenden wir zunächst m-mal die Produktion $S \rightarrow aSb$ an, gefolgt von der Produktion $S \rightarrow A$ und dann noch $n-m-1$ mal die Produktion $A \rightarrow aA$ und einmal die Produktion $A \rightarrow a$:

$$S \overset{*}{\Longrightarrow} a^m Sb^m \overset{*}{\Longrightarrow} a^m Ab^m \overset{*}{\Longrightarrow} a^m a^{n-m-1} Ab^m \overset{*}{\Longrightarrow} a^m a^{n-m-1} ab^m = a^n b^m$$

   Damit gilt auch $w \in L(G)$.

   Sei andersherum $w \in L(G)$, dann gilt zunächst $w \in \{a\}^+$ für jedes w mit $A \overset{*}{\Longrightarrow} w$. In einer Ableitung eines Wortes w kann nun zunächst i mal die Produktion $S \rightarrow aSb$ benutzt werden, was zu $a^i Sb^i$ führt. Um ein Wort zu generieren, dass nur aus Terminalzeichen besteht, muss nun irgendwann die Produktion $S \rightarrow A$ benutzt werden. Abschließend kann wie oben erwähnt A nur zu einem a^j mit $j \geq 1$ abgeleitet werden. Insgesamt haben wir so das Wort $a^{i+j}b^i$, was wegen $i + j > i$ in M ist.

Übungsaufgabe 4.4: Bestimmen Sie zu der folgenden Grammatik G eine äquivalente Grammatik in (ggf. erweiterter) Chomsky-Normalform *nach dem Verfahren aus der Vorlesung (siehe auch Theorem 9.22 im Skript - beachten Sie aber die dort falsche Reihenfolge der Schritte 2 und 3)*. Stellen Sie die einzelnen Schritte des Verfahrens deutlich und gut nachvollziehbar dar. Geben Sie dazu auch (wo nötig) die Produktionen der neu entstehenden Grammatiken an.

von 6

$$
\begin{array}{rcl}
S & \longrightarrow & AB \mid AB1B \mid AD \mid DE \mid C \\
A & \longrightarrow & EBEE \\
B & \longrightarrow & A \mid E \mid SE \\
C & \longrightarrow & \lambda \\
D & \longrightarrow & CBG \mid 0G \mid 1D \\
E & \longrightarrow & 0 \mid 1 \\
F & \longrightarrow & 0E \mid EF \mid CC
\end{array}
$$

Lösung: Wir wenden das sechs-schrittige Verfahren aus dem Skript (Theorem 9.22) an, allerdings mit der Korrektur aus der Vorlesung, dass die Schritte 2 und 3 vertauscht werden müssen.

1. Schritt. Im ersten Schritt bestimmen wir (nach Theorem 9.12) die Menge der Nonterminale, die auf λ ableiten zu $V_\lambda = \{C, F, S\}$ (mit $M_0 = \{C\}$, $M_1 = \{C, F, S\}$ und $M_2 = \{C, F, S\} = M_1 = V_\lambda$). Dies führt zu

$$
\begin{array}{rcl}
S & \longrightarrow & AB \mid AB1B \mid AD \mid DE \mid C \\
A & \longrightarrow & EBEE \\
B & \longrightarrow & A \mid E \mid SE \\
D & \longrightarrow & CBG \mid 0G \mid 1D \mid BG \\
E & \longrightarrow & 0 \mid 1 \\
F & \longrightarrow & 0E \mid EF \mid CC \mid E \mid C
\end{array}
$$

(Man beachte, dass durch Streichen von S in der Produktion $B \longrightarrow SE$ die Produktion $B \longrightarrow E$ ensteht, die aber bereits enthalten ist.)

2. Schritt. Als Kettenregeln sind $S \longrightarrow C$, $B \longrightarrow A$, $B \longrightarrow E$, $F \longrightarrow E$ und $F \longrightarrow C$ enthalten, so dass wir als Produktion einer Grammatik ohne Kettenregeln folgendes erhalten:

$$
\begin{array}{rcl}
S & \longrightarrow & AB \mid AB1B \mid AD \mid DE \\
A & \longrightarrow & EBEE \\
B & \longrightarrow & SE \mid EBEE \mid 0 \mid 1 \\
D & \longrightarrow & CBG \mid 0G \mid 1D \mid BG \\
E & \longrightarrow & 0 \mid 1 \\
F & \longrightarrow & 0E \mid EF \mid CC \mid 0 \mid 1
\end{array}
$$

Bspw. wurde die Ableitung $B \Rightarrow A \Rightarrow EBEE$ durch die Produktion $B \longrightarrow EBEE$ ersetzt, die nun die gleiche Ableitung erlaubt, aber ohne den Zwischenschritt über die Kettenregeln $B \longrightarrow A$.

3. Schritt. Im dritten Schritt wird die Grammatik reduziert. Zunächst werden die produktiven Symbole bestimmt und die Produktionen entsprechend eingeschränkt. Wir erhalten:

$$
\begin{array}{rcl}
M_0 & = & \{0, 1\} \\
M_1 & = & \{0, 1, B, E, F\} \\
M_2 & = & \{0, 1, B, E, F, A\} \\
M_3 & = & \{0, 1, B, E, F, A, S\}
\end{array}
$$

Wegen $M_3 = M_4$ stoppt das Verfahren. Wir erhalten nun eine Grammatik mit den folgenden Produktionen:

$$\begin{array}{rcl} S & \longrightarrow & AB \mid AB1B \\ A & \longrightarrow & EBEE \\ B & \longrightarrow & SE \mid EBEE \mid 0 \mid 1 \\ E & \longrightarrow & 0 \mid 1 \\ F & \longrightarrow & 0E \mid EF \mid 0 \mid 1 \end{array}$$

Die erreichbaren Symbole bestimmen wir nun zu:

$$\begin{array}{rcl} M_0 & = & \{S\} \\ M_1 & = & \{S, A, B\} \\ M_2 & = & \{S, A, B, E\} \end{array}$$

Wegen $M_2 = M_3$ stoppt das Verfahren wieder und wir erhalten

$$\begin{array}{rcl} S & \longrightarrow & AB \mid AB1B \\ A & \longrightarrow & EBEE \\ B & \longrightarrow & SE \mid EBEE \mid 0 \mid 1 \\ E & \longrightarrow & 0 \mid 1 \end{array}$$

4. Schritt. Im vierten Schritt werden nun lediglich Terminalzeichen, die auf rechten Seiten auftreten, deren Länge größer als 1 ist, durch Nonterminale ersetzt. Dies tritt hier nur in der Produktion $S \longrightarrow AB1B$ auf. Wir erhalten so

$$\begin{array}{rcl} S & \longrightarrow & AB \mid AB\langle 1\rangle B \\ A & \longrightarrow & EBEE \\ B & \longrightarrow & SE \mid EBEE \mid 0 \mid 1 \\ E & \longrightarrow & 0 \mid 1 \\ \langle 1\rangle & \longrightarrow & 1 \end{array}$$

Man beachte, dass $\langle 1\rangle$ *ein* Nonterminalsymbol ist.

5. Schritt. Im fünften Schritt werden zu lange rechte Seiten einer Produktion verkürzt. Dies geschieht, indem stets das am weitesten rechts stehende Symbol beibehalten wird und die Symbole davor zu einem neuen Nonterminal zusammengefasst werden. Aus der Produktion $A \longrightarrow EBEE$ wird so $A \longrightarrow (EBE)E$, wobei (EBE) ein neues Nonterminalzeichen ist. Ferner werden die Produktionen $(EBE) \longrightarrow (EB)E$ und $(EB) \longrightarrow EB$ hinzugefügt. Man beachte, dass nach Konstruktion diese Produktionen erzeugt werden und nicht etwa $A \longrightarrow (EB)(EE)$ usw. Insgesamt erhält man so

$$\begin{array}{rcl} S & \longrightarrow & AB \mid (AB\langle 1\rangle)B \\ A & \longrightarrow & (EBE)E \\ B & \longrightarrow & SE \mid (EBE)E \mid 0 \mid 1 \\ E & \longrightarrow & 0 \mid 1 \\ \langle 1\rangle & \longrightarrow & 1 \\ (AB\langle 1\rangle) & \longrightarrow & (AB)\langle 1\rangle \\ (AB) & \longrightarrow & AB \\ (EBE) & \longrightarrow & (EB)E \\ (EB) & \longrightarrow & EB \end{array}$$

wobei $(AB\langle 1\rangle)$, (AB), (EBE) und (EB) neue Nonterminalsymbole sind.

6. Schritt. Da im ersten Schritt $S \in V_\lambda$ ermittelt wurde, gilt für die bisher erzeugte Grammatik G' und die ursprüngliche Grammatik G die Beziehung $L(G)' = L(G) \setminus \{\lambda\}$ und $\lambda \in L(G)$, d.h. G'

ist nicht äquivalent zu G. Um dies nun zu erreichen, wird ein neues Startsymbol S_{neu} eingeführt sowie die folgenden Produktionen

$$S_{neu} \quad \longrightarrow \quad \lambda \mid AB \mid (AB\langle 1\rangle)B$$

alle anderen Produktionen bleiben erhalten (insb. sind auch jene mit S auf der linken Seite weiterhin nötig, da S in der Regel $B \longrightarrow SE$ auf der rechten Seite auftritt).

Damit haben wir nun eine Grammatik G'' mit $L(G'') = L(G)$ und G'' ist in erweiterter Chomsky-Normalform.

von
2

Übungsaufgabe 4.5: Nutzen Sie das Pumping Lemma für reguläre Sprachen, um zu zeigen, dass

$$L := \{a^p \mid \ p \text{ ist eine Primzahl}\}$$

keine reguläre Sprache ist. Ist L kontextfrei?

Lösung: Angenommen L wäre regulär. Dann gilt das Pumping Lemma. Sei n die Zahl aus dem Pumping Lemma. Wir können nun nicht einfach das Wort a^n betrachten, da dies zwar eine erlaubte Länge hat, aber nicht zwingend in L ist. Sei daher $p \geq n$ die kleinste Primzahl größer oder gleich n (nachfolgendes klappt mit einer beliebigen Primzahl, die größer oder gleich n ist). Wir betrachten $z = a^p$. Es ist $|z| \geq n$ und $z \in L$, daher muss es nach dem Pumping Lemma eine Zerlegung von z in uvw geben, die die Bedingungen i, ii und iii des Pumping Lemmas erfüllt. Wir betrachten nur Zerlegungen, die i und ii erfüllen (die anderen sind ja schon im Widerspruch zu einer der drei Eigenschaften, nämlich zu i oder ii) und zeigen, dass wir zu einem Widerspruch zu iii gelangen können.

Sei also $z = uvw$ eine Zerlegung mit $|uv| \leq n$ und $|v| \geq 1$. Es gibt dann ein j mit $v = a^j$. Die Frage ist nun, wie wir "pumpen" müssen, um $uv^iw = a^{p+(i-1)\cdot j} \notin L$ zu erreichen, denn dann haben wir den gewünschten Widerspruch (man beachte, dass $uv^1w = a^p$ ist. Daher ist $uv^iw = uv^{1+(i-1)}w = a^{p+(i-1)\cdot j}$). Die Zahl $p + (i-1) \cdot j$ darf hierfür keine Primzahl sein. Wir dürfen, um dies zu erreichen, $i \geq 0$ frei wählen. Setzen wir $i := p+1$ so haben wir den gewünschten Widerspruch, denn $p+(i-1)\cdot j = p+(p+1-1)\cdot j = p+p\cdot j = p\cdot(1+j)$ ist eine aus p und $(1+j) \geq 2$ zusammengesetzte Zahl und damit keine Primzahl. Folglich ist $uv^{p+1}w = a^{p+(p+1-1)\cdot j} = a^{p+p\cdot j} = a^{p\cdot(1+j)}$ nicht in L_2 und wir haben den gewünschten Widerspruch.

Die Frage, ob L kontextfrei ist, kann man mit dem Pumping Lemma für kontextfreie Sprachen ebenfalls negativ beantworten. Der Beweis verläuft sehr ähnlich, da jedes Wort in L nur aus as besteht. Der Unterschied ist lediglich, dass nun $|vwx| \leq n$ und $|vx| \geq 1$ gilt. Man pumpt aber wie eben dann $p + 1$ mal an v und x (betrachtet also $uv^{p+1}wx^{p+1}y$) und gelangt so zu dem gleichen Widerspruch. L ist also nicht einmal kontextfrei.