
Einführung in die Theoretische Informatik

Hausaufgabe 1 (5 Punkte)

Wir betrachten die Grammatik $G = (\{S, T\}, \{a, b\}, P, S)$ mit den Produktionen

$$\begin{aligned} S &\rightarrow aSb \mid SS \mid \epsilon, \\ T &\rightarrow TaTb \mid \epsilon. \end{aligned}$$

Zeigen Sie jeweils per Induktion:

1. $L(T) \subseteq L(S)$.
2. Wenn $w \in L(T)$, dann ist auch $awb \in L(T)$.
3. Wenn $v \in L(T)$ und $w \in L(T)$, dann ist auch $vw \in L(T)$.
4. $L(S) \subseteq L(T)$.

Lösungsvorschlag

1. Sei $w \in L(T)$. Mit Induktion über die Erzeugung von w zeigen wir, dass $w \in L(S)$.
Dafür müssen wir alle T -Regeln untersuchen:

$T \rightarrow \epsilon$: Wenn $w = \epsilon$ dann ist es auch (mit der Regel $S \rightarrow \epsilon$) aus S ableitbar, und damit in $L(S)$.

$T \rightarrow TaTb$: Wir nehmen an, dass $w = uavb$ mit $u, v \in L(T)$. Nach Induktionshypothese sind damit u und v auch in $L(S)$. Damit gibt es die Ableitung $S \rightarrow SS \rightarrow SaSb \rightarrow^* uaSb \rightarrow^* uavb = w$, und somit ist $w \in L(S)$.

2. Wieder Induktion über die Erzeugung von $w \in L(T)$:

$T \rightarrow \epsilon$: Wenn $w = \epsilon$ dann ist $awb = ab \in L(T)$, mit der Ableitung $T \rightarrow TaTb \rightarrow^2 \epsilon a \epsilon b = ab$.

$T \rightarrow TaTb$: Wir nehmen an, dass $w = uavb$ und $u, v \in L(T)$. Dann ist auch $auavbb \in L(T)$, mit der Ableitung $T \rightarrow TaTb \rightarrow aTb \rightarrow aTaTbb \rightarrow^* auavbb$. (Man beachte, dass hier die Induktionshypothese nicht verwendet wurde, womit die Induktion zu einer einfachen Fallunterscheidung degeneriert.)

3. Induktion über die Erzeugung von w :

$T \rightarrow \epsilon$: Wenn $w = \epsilon$ dann ist $vw = v$ und trivialerweise in $L(T)$.

$T \rightarrow TaTb$: Wir nehmen an, dass $w = uaw'b$ und $u, w' \in L(T)$. Nach Induktionshypothese ist dann auch $vu \in L(T)$. Wir zeigen, dass $vuaw'b \in L(T)$ mit der folgenden Ableitung: $T \rightarrow TaTb \rightarrow^* vuaTb \rightarrow^* vuaw'b$.

4. Sei $w \in L(S)$. Wir zeigen $w \in L(T)$ mit Induktion über die Erzeugung von w aus S . Dabei betrachten wir die Regeln für S :
- $S \rightarrow aSb$: Sei $w = avb$ mit $v \in L(S)$. Nach Induktionshypothese ist damit $v \in L(T)$ und mit Aufgabenteil 2 damit auch $avb \in L(T)$.
- $S \rightarrow SS$: Sei $w = uv$ mit $u, v \in L(S)$. Nach Induktionshypothese sind damit auch $u, v \in L(T)$ und mit Aufgabenteil 3 gilt damit auch $uv \in L(T)$.
- $S \rightarrow \epsilon$: Dieser Fall ist klar, da $T \rightarrow \epsilon$ direkt ableitbar ist.

Hausaufgabe 2 (5 Punkte)

Wir betrachten die Grammatik $G = (\{S\}, \{a, b\}, P, S)$ mit den folgenden Produktionen:

$$S \rightarrow aSbSb \mid SS \mid \epsilon.$$

1. Zeigen Sie induktiv, dass für alle ableitbaren Wörter $w \in L(G)$ die Anzahl der enthaltenen b doppelt so groß ist wie die Anzahl der enthaltenen a , d. h., dass gilt $\#_b(w) = 2 \cdot \#_a(w)$.
2. Zeigen Sie, dass alle Satzformen $\alpha_n = (ab)^n Sb^n$ für $n \geq 0$ in G ableitbar sind.

Lösungsvorschlag

1. Induktion über die Erzeugung von $w \in L(S)$:
 $S \rightarrow aSbSb$: Sei $w = aubvb$ mit $u, v \in L(S)$. Nach Induktionshypothese gilt die Behauptung für u und v , und damit ist auch

$$\#_b(w) = \#_b(u) + \#_b(v) + 2 = 2 \cdot (\#_a(u) + \#_a(v) + 1) = 2 \cdot \#_a(w).$$

Der Fälle $S \rightarrow SS$ und $S \rightarrow \epsilon$ sind analog.

2. Induktion über n .
 $n = 0$: Da S das Startsymbol ist, gilt $S \xrightarrow{0}_G S = (ab)^0 Sb^0$.
 $n \rightsquigarrow n + 1$: Nach Induktionshypothese gilt $S \xrightarrow{*}_G (ab)^n Sb^n$. Dann gilt auch $S \xrightarrow{*}_G (ab)^n Sb^n \xrightarrow{G} (ab)^n aSbSbb^n \xrightarrow{G} (ab)^n abSbb^n = (ab)^{n+1} Sb^{n+1}$.

Hausaufgabe 3 (5 Punkte)

Sei $\Sigma = \{0, 1, (,), \cup, *, \emptyset, \epsilon\}$ die Zeichenmenge, aus der reguläre Ausdrücke über dem Alphabet $\{0, 1\}$ gebildet werden. (Wir schreiben hier \cup anstelle von $|$, um Zeichenverwirrung beim Aufschreiben der Grammatik zu vermeiden.)

1. Geben Sie eine *eindeutige* kontextfreie Grammatik an, die die Menge der regulären Ausdrücke beschreibt. Die Grammatik soll so konstruiert sein, dass ungeklammerte Ausdrücke stets als rechts geklammert geparsert werden (z.B. 010 als 0(10)) und die Prioritäten der Operatoren berücksichtigt sind.
2. Geben Sie einen Syntaxbaum für das Wort $01^*0|1$ an.
*(Korrektur: Nach obiger Schreibweise muss hier natürlich $01^*0 \cup 1$ stehen)*

Lösungsvorschlag

- Die Menge der regulären Ausdrücke ist bereits induktiv definiert, und eine solche Definition kann man als Grammatik $G' = (\{R\}, \Sigma, P, R)$ mit folgenden Produktionen auffassen:

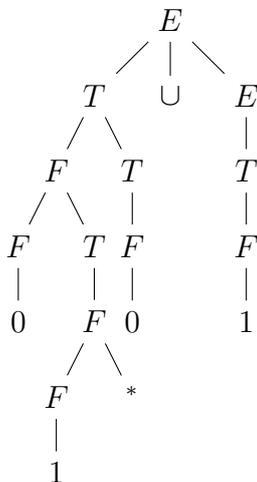
$$R \rightarrow 0 \mid 1 \mid \epsilon \mid \emptyset \mid RR \mid R \cup R \mid R^* \mid (R)$$

Diese Grammatik ist allerdings nicht eindeutig, da z.B. der Ausdruck $0 \cup 00$ zwei Syntaxbäume hat, die jeweils der Klammerung $0 \cup (00)$ bzw. $(0 \cup 0)0$ entsprechen. Eine eindeutige Grammatik erhält man, indem man mehrere Nichtterminale verwendet ähnlich wie in Beispiel 3.3 aus der Vorlesung. Wir setzen $G = (\{E, T, F\}, \Sigma, P, E)$ mit den Produktionen

$$\begin{aligned} E &\rightarrow T \mid T \cup E \\ T &\rightarrow F \mid FT \\ F &\rightarrow 0 \mid 1 \mid \epsilon \mid \emptyset \mid (E) \mid F^* \end{aligned}$$

Begründung der Eindeutigkeit: Jedes Wort w in $L(G)$ ist entweder eine Vereinigung oder es ist keine Vereinigung. Es steht also fest, ob die erste angewandte Produktion $E \rightarrow T \cup E$ oder $E \rightarrow T$ ist, und nur eine der beiden Produktionen kann als erstes angewandt werden. Entsprechendes gilt für $w \in L(T)$ und die Konkatenation. Bei allen Ableitungen von Wörtern $w \in L(F)$ aus der Variablen F ist ebenfalls die erste Produktionsanwendung durch w vorgegeben.

2.



Hausaufgabe 4 (5 Punkte)

Wir betrachten *erweiterte* kontextfreie Grammatiken, bei denen auf der rechten Seite jeder Produktion ein regulärer Ausdruck α über $\Sigma \cup V$ anstelle einer einzelnen Satzform w über $\Sigma \cup V$ stehen darf. Beim Anwenden der Produktion darf dann eine beliebige Satzform aus $L(\alpha)$ eingesetzt werden.

Diese Art von Grammatik ist in der Praxis oft praktischer, da sich mit Hilfe der Sternoperation Wiederholung leichter ausdrücken lässt. Ein Beispiel ist die Grammatik

$$G = (\{0, 1, ;, [,], \{S\}, P, S\})$$

mit den Produktionen

$$S \rightarrow 1(0 \mid 1)^* \mid [S(;S)^*]$$

1. Zeigen Sie, dass erweiterte kontextfreie Grammatiken nicht ausdrucksstärker sind als gewöhnliche kontextfreie Grammatiken, indem Sie ein Verfahren angeben, wie eine erweiterte Grammatik in eine gewöhnliche Grammatik transformiert werden kann.
2. Wenden Sie Ihr Verfahren auf die oben angegebene Grammatik G an.

Lösungsvorschlag

1. Die erweiterte Grammatik wird schrittweise in eine gewöhnliche Grammatik transformiert, indem man Produktionen mit einem regulären Ausdruck auf der rechten Seite in mehrere einfachere Produktionen zerlegt. Dies zeigen wir anhand einer Produktion $A \rightarrow \alpha$ für die verschiedenen Arten, wie α gebildet sein kann:

$\alpha \in \Sigma \cup \{\epsilon\}$: Hier ist nichts zu tun, da die Produktion schon die richtige Form hat. (Strenggenommen ersetzt man $A \rightarrow \epsilon$ durch $A \rightarrow \epsilon$.)

$\alpha = \emptyset$: Hier kann die Produktion einfach entfernt werden.

$\alpha = \alpha_1\alpha_2$: Man führt neue Nichtterminale A_1 und A_2 ein und ersetzt die Produktion durch die drei neuen Produktionen $A \rightarrow A_1A_2$, $A_1 \rightarrow \alpha_1$, $A_2 \rightarrow \alpha_2$.

$\alpha = (\alpha_1 \mid \alpha_2)$: Man ersetzt die Produktion $A \rightarrow (\alpha_1 \mid \alpha_2)$, wobei \mid die Operation auf regulären Ausdrücken bezeichnet, einfach durch zwei Produktionen $A \rightarrow \alpha_1$ und $A \rightarrow \alpha_2$. In der Kurznotation $A \rightarrow \alpha_1 \mid \alpha_2$ sieht das Ergebnis fast so aus wie vorher, allerdings handelt es sich nun um zwei getrennte Produktionen.

$\alpha = \alpha_1^*$: Man führt ein neues Nichtterminale A_1 ein und ersetzt die Produktion durch $A \rightarrow \epsilon$, $A \rightarrow A_1A$, $A_1 \rightarrow \alpha_1$.

2. Wir starten mit der gegebenen Grammatik, die wir einfach als Menge von Produktionen angeben:

$$S \rightarrow 1(0 \mid 1)^* \mid [S(;S)^*]$$

Am konkreten Beispiel weichen wir vom allgemeinen Verfahren ab, wenn dadurch nur unnütze Schreiarbeit entstehen würde. Zum Beispiel zerlegen wir die erste Produktion wie folgt, und verzichten auf die Einführung eines Nichtterminals für 1:

$$S \rightarrow 1A \mid [S(;S)^*] \qquad A \rightarrow (0 \mid 1)^*$$

Dann weiter:

$$S \rightarrow 1A \mid [S(;S)^*] \qquad A \rightarrow \epsilon \mid BA \qquad B \rightarrow 0 \mid 1$$

Bei der zweiten Produktion für S kürzen wir ebenfalls etwas ab, indem wir nur eine neue Variable C für den Sternausdruck einführen:

$$S \rightarrow 1A \mid [SC] \qquad A \rightarrow \epsilon \mid BA \qquad B \rightarrow 0 \mid 1 \\ C \rightarrow (;S)^*$$

Dieser Ausdruck wird nun noch wie oben aufgelöst, womit wir eine gewöhnliche kontextfreie Grammatik erhalten:

$$\begin{array}{lll} S \rightarrow 1A \mid [SC] & A \rightarrow \epsilon \mid BA & B \rightarrow 0 \mid 1 \\ C \rightarrow \epsilon \mid DC & D \rightarrow ;S & \end{array}$$

Vorbereitung 1

Wandeln Sie die folgende Grammatik in Chomsky-Normalform um:

$$\begin{aligned} S &\rightarrow 0A0 \mid 1B1 \mid BB, \\ A &\rightarrow D, \quad B \rightarrow S \mid A, \\ C &\rightarrow S \mid \epsilon, \quad D \rightarrow C. \end{aligned}$$

Lösungsvorschlag

Wir können Variablen mit nur einer einzigen Produktion durch "Einsetzen" eliminieren. Somit vereinfachen wir die Grammatik zunächst zu

$$S \rightarrow 0A0 \mid 1B1 \mid BB, \quad A \rightarrow C, \quad B \rightarrow S \mid A, \quad C \rightarrow S \mid \epsilon.$$

und schließlich zu

$$S \rightarrow 0C0 \mid 1B1 \mid BB, \quad B \rightarrow S \mid C, \quad C \rightarrow S \mid \epsilon.$$

Nun eliminieren wir die ϵ -Produktionen. Dazu bauen wir die Menge \hat{P} wie in Lemma 3.26. Wir müssen nacheinander die Produktionen $S \rightarrow 00, B \rightarrow \epsilon, S \rightarrow 11, S \rightarrow B, S \rightarrow \epsilon$ hinzufügen. Anschliessend können wir alle ϵ -Produktionen entfernen und erhalten

$$S \rightarrow 0C0 \mid 00 \mid 1B1 \mid 11 \mid BB \mid B, \quad B \rightarrow S \mid C, \quad C \rightarrow S.$$

Wir eliminieren jetzt die Kettenproduktionen, und zwar wieder durch Einsetzen, da die betreffenden Variablen nur eine einzige Produktion haben. Wir beginnen mit $C \rightarrow S$

$$S \rightarrow 0S0 \mid 00 \mid 1B1 \mid 11 \mid BB \mid B, \quad B \rightarrow S$$

... dann folgt $B \rightarrow S$

$$S \rightarrow 0S0 \mid 00 \mid 1S1 \mid 11 \mid SS \mid S$$

... und zuletzt $S \rightarrow S$, die einfach entfernt werden kann:

$$S \rightarrow 0S0 \mid 00 \mid 1S1 \mid 11 \mid SS$$

Nun führen wir neue Variablen N und E für 0 bzw. 1 ein:

$$S \rightarrow NSN \mid NN \mid ESE \mid EE \mid SS, \quad N \rightarrow 0, \quad E \rightarrow 1$$

Die Dreier-Produktionen eliminieren wir, indem wir Zwischensymbole T_1 und T_2 einfügen:

$$S \rightarrow NT_1 \mid NN \mid ET_2 \mid EE \mid SS, \quad T_1 \rightarrow SN, \quad T_2 \rightarrow SE, \quad N \rightarrow 0, \quad E \rightarrow 1$$

Diese Grammatik ist nun in Chomsky-Normalform und beschreibt dieselbe Sprache, ausgenommen dem ϵ .

Vorbereitung 2

Welche Symbole in der folgenden Grammatik sind erzeugend, welche erreichbar, und welche nützlich?

$$\begin{aligned} S &\rightarrow A \mid a, \\ A &\rightarrow BC, \quad B \rightarrow AE \mid C, \\ C &\rightarrow aS, \quad D \rightarrow aB \mid FS \\ E &\rightarrow EA, \quad F \rightarrow Ca \mid a. \end{aligned}$$

Lösungsvorschlag

Alle Symbole außer E sind erzeugend. Alle Symbole außer D und F sind erreichbar. Die anderen Symbole sind nützlich, wie die folgende Ableitung zeigt:

$$S \rightarrow A \rightarrow BC \rightarrow CC \xrightarrow{2} aSaS \xrightarrow{2} aaaa$$

Tutoraufgabe 1

Wir betrachten die kontextfreie Grammatik $G = (V, \{a, b\}, P, S)$ mit Produktionen in Chomsky-Normalform

$$\begin{aligned} S &\rightarrow YC_3 \mid YX \mid AC_3 \mid AX \mid YY \mid YB \mid YA \mid AY \mid AB \mid AA, \\ X &\rightarrow ZC_2 \mid ZB \mid BZ \mid BB, \\ Y &\rightarrow AC_1 \mid AY \mid AA, \\ C_1 &\rightarrow YZ \mid AZ, \\ C_2 &\rightarrow BZ, \\ Z &\rightarrow YC_3 \mid YX \mid AC_3 \mid AX \mid YY \mid YB \mid YA \mid AY \mid AB \mid AA, \\ C_3 &\rightarrow XZ \mid YZ \mid BZ \mid AZ, \\ A &\rightarrow a, \\ B &\rightarrow b. \end{aligned}$$

Entscheiden Sie mit Hilfe des Cocke-Younger-Kasami-Algorithmus, ob die Wörter $aaaa$ und $abab$ in der Sprache $L(G)$ enthalten sind! Geben Sie gegebenenfalls Ableitungen an!

Lösungsvorschlag

Zur Entscheidung der Ableitbarkeit eines Wortes $w = a_1a_2 \dots a_n \in \{a, b\}^+$ berechnet der CYK-Algorithmus die Mengen $V_{i,j} = \{A \in V \mid A \xrightarrow{*} a_i \dots a_j\}$ wie folgt.

$$\begin{aligned} V_{i,i} &= \{A \in V \mid A \rightarrow a_i \in P\} \\ \text{und für } i < j \quad V_{i,j} &= \bigcup_{i \leq k < j} \{A \in V \mid (A \rightarrow BC) \in P \wedge B \in V_{i,k} \wedge C \in V_{k+1,j}\}. \end{aligned}$$

Die Frage ist dann $S \in V_{1,n}$?

Berechnungstabelle für die Mengen V_{ij} mit $w = aabaa$:

¹⁵ S,Z,C ₃				
¹⁴ ∅	²⁵ S,Z			
¹³ S,Z,X,C ₁ ,C ₃	²⁴ ∅	³⁵ X,C ₁ ,C ₃		
¹² S,Y,Z	²³ S,Z	³⁴ ∅	⁴⁵ S,Y,Z	
¹¹ A	²² A	³³ B	⁴⁴ A	⁵⁵ A
a	a	b	a	a

Das Wort $aaaa$ ist also in der Sprache enthalten.

Ableitung:

$$\begin{aligned} S(1, 5) &\rightarrow Y(1, 2)X(3, 5) \rightarrow A(1, 1)A(2, 2)X(3, 5) \rightarrow AAB(3, 3)Z(4, 5) \\ &\rightarrow AABA(4, 4)A(5, 5) \xrightarrow{G^*} aabaa. \end{aligned}$$

Die Indizes bezeichnen die Menge $V_{i,j}$, in der die betreffende Variable berechnet wurde, und sind natürlich nicht Teil der Sprache.

Berechnungstabelle für $w = abab$:

¹⁴ S, Z			
¹³ \emptyset	²⁴ X, C ₂ , C ₃		
¹² S, Z	²³ \emptyset	³⁴ S, Z	
¹¹ A	²² B	³³ A	⁴⁴ B
a	b	a	b

Das Wort $abab$ ist also auch in der Sprache enthalten.

Ableitung: $S \rightarrow A(1, 1)X(2, 4) \rightarrow AB(2, 2)Z(3, 4) \rightarrow ABA(3, 3)B(4, 4) \rightarrow_G^* abab$.

Tutoraufgabe 2

Skizzieren Sie ein Verfahren, welches für eine gegebene kontextfreie Grammatik G entscheidet, ob $L(G)$ endlich ist.

Lösungsvorschlag

Konzeptionell am einfachsten ist es, anzunehmen, dass die Grammatik in CNF vorliegt und nur noch nützliche Symbole enthält. Dann genügt es, die Grammatik auf Zyklen zu testen: Ein Zyklus ist dabei ein Nichtterminal A , so dass $A \rightarrow_G^* \alpha A \beta$.

Tutoraufgabe 3

Gegeben sei die Sprache $L = \{a^n b^{4n} \mid n \in \mathbb{N}\}$.

1. Geben Sie eine kontextfreie Grammatik G an, die L erzeugt.
2. Konstruieren Sie mit Standardverfahren aus der Vorlesung zu G einen Kellerautomaten K , der L durch Endzustände akzeptiert.

Lösungsvorschlag

1. G ergibt sich analog zur bekannten Grammatik für $\{a^n b^n \mid n \in \mathbb{N}\}$ wie folgt.

$$\begin{aligned} S &\rightarrow aSB \mid \epsilon, \\ B &\rightarrow bbbb. \end{aligned}$$

2. Wir gehen aus von der Grammatik G der Teilaufgabe 1 und bringen die Produktionen gemäß Satz 3.56 der Vorlesung in die Form $A \rightarrow aA_1A_2 \dots A_k$ mit $a \in \Sigma \cup \{\epsilon\}$. Die transformierte Grammatik G_t erzeugt die gleiche Sprache wie G , also L .

Im zweiten Schritt definieren wir kanonisch wie in Satz 3.56 einen PDA A , der L bzw. $L(G_t)$ mit leerem Keller akzeptiert.

Im dritten Schritt konstruieren wir dann nach dem Verfahren des Satzes 3.53 zu A einen PDA A' , der L mit Endzuständen akzeptiert.

Grundsätzlich wird zunächst für jedes Zeichen $x \in \Sigma$ eine neue Variable A_x zusammen mit einer Produktion $A_x \rightarrow x$ eingeführt. Man sieht dem folgenden transformierten Produktionensystem aber sofort an, dass nur für b eine neue Variable eingeführt zu werden braucht.

Das transformierte Produktionensystem lautet

$$\begin{aligned} S &\rightarrow aSB \mid \epsilon, \\ B &\rightarrow bA_bA_bA_b, \\ A_b &\rightarrow b. \end{aligned}$$

Der PDA A ist dann definiert durch $A = (\{q_0\}, \Sigma, \{S, B, A_b\}, q_0, S, \delta)$ mit

$$\delta(q_0, x, Z) = \{(q_0, \beta) \mid Z \rightarrow x\beta\}$$

für alle $x \in \Sigma \cup \{\epsilon\}$ und $Z \in \{S, B, A_b\}$.

Im letzten Schritt setzen wir nach Satz 3.53

$$A' = (\{q_0, q'_0, q_f\}, \Sigma, \{S, B, A_b, Z'_0\}, q'_0, Z'_0, \delta', \{q_f\})$$

mit

$$\begin{aligned} \delta'(q'_0, \epsilon, Z'_0) &= \{(q_0, SZ'_0)\}, \\ \delta'(q_0, x, Z) &= \{(q_0, \beta) \mid Z \rightarrow x\beta\} \quad \text{für } x \in \Sigma \cup \{\epsilon\}, Z \in \{S, B, A_b\}, \\ \delta'(q_0, \epsilon, Z'_0) &= \{(q_f, Z'_0)\}. \end{aligned}$$