Technische Universität München
Fakultät für Informatik
Prof. Tobias Nipkow, Ph.D.
Sascha Böhme, Lars Noschinski

Sommersemester 2011
Lösungsblatt 5
6. Juni 2011

Einführung in die Theoretische Informatik

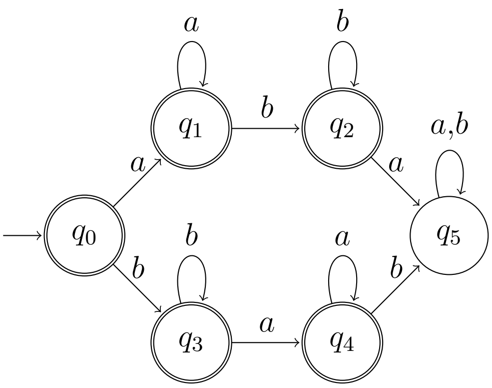
***Hinweis:** Bitte beachten Sie unbedingt die Hinweise zum Übungsablauf und zu den Aufgabentypen auf der THEO-Website (`http://theo.in.tum.de/`).*

Hausaufgabe 1 (5 Punkte)

Betrachten Sie die Sprache $L = L(a^*b^* \mid b^*a^*)$ über $\Sigma = \{a, b\}$. Geben Sie dazu einen minimalen DFA an und verwenden Sie diesen, um alle Äquivalenzklassen von $\equiv_L$ zu bestimmen und jede durch einen regulären Ausdruck zu beschreiben.

Lösungsvorschlag

Die Sprache L wird durch folgenden DFA A akzeptiert:



	0	1	2	3	4
1	$2/b$				
2	$1/a$	$1/a$			
3	$2/a$	$2/a$	$1/a$		
4	$1/b$	$1/b$	$1/a$	$1/b$	
5	X	X	X	X	X

Wie man an der nebenstehenden Tabelle sehen kann, sind keine zwei Zustände äquivalent, also ist der Automat minimal. Bezeichne δ die Übergangsrelation unseres Automaten A. Wir wissen daher: Die Äquivalenzklassen der Sprache entsprechen den Zuständen des Automaten. Formal: Für zwei Wörter $u, v \in \Sigma^*$ gilt:

$$[u] \equiv_L [v] \Leftrightarrow \hat{\delta}(q_0, u) \equiv_A \hat{\delta}(q_0, v) \Leftrightarrow \hat{\delta}(q_0, u) = \hat{\delta}(q_0, v)$$

Der letzte Schritt gilt, da A minimal ist.
Die Sprachäquivalenzklasse zum Zustand q_i entspricht also genau der Sprache $L(A_i)$ zum Automaten $A_i = (\{q_0, \cdots, q_5\}, \Sigma, \delta, \{q_i\})$. Für diese Sprachen können wir leicht reguläre Ausdrücke bestimmen:

$$[\epsilon] : \boldsymbol{\epsilon} \qquad [a] : a^+ \qquad [b] : b^+ \qquad [ab] : a^+b^+ \qquad [ba] : b^+a^+ \qquad [aba] : \Sigma^*(ab^+a \mid ba^+b)\Sigma^*$$

sind die Äquivalenzklassen von $\equiv_L$.

Hausaufgabe 2 (5 Punkte)

Wir betrachten die Sprache $L = L(a^*b^*c^*) \setminus \{a^nb^nc^n \mid n \geq 0\}$ über dem Alphabet $\Sigma = \{a, b, c\}$.

1. Zeigen Sie, dass L kontextfrei ist, indem Sie eine kontextfreie Grammatik G für diese Sprache angeben. Ein Beweis, dass $L(G) = L$ ist, wird nicht erwartet.

2. Geben Sie je eine G-Ableitung für die Wörter *abbcc* und *abccc* an.

Lösungsvorschlag

1. Eine mögliche Grammatik ist $G = (\{S, T_{ab}, T_{ac}, T_{bc}, A, B, C\}, \Sigma, P, S\}$, wobei P aus den folgenden Produktionen besteht:

$$\begin{aligned} S &\to T_{ab}C \mid T_{ac} \mid AT_{bc} \\ T_{ab} &\to aT_{ab}b \mid aA \mid bB \\ T_{ac} &\to aT_{ac}c \mid aAB \mid BcC \\ T_{bc} &\to bT_{bc}c \mid bB \mid cC \\ A &\to aA \mid \epsilon \\ B &\to bB \mid \epsilon \\ C &\to cC \mid \epsilon \end{aligned}$$

   Die drei S-Produktionen unterscheiden dabei die Sprachen, in denen a und b bzw. a und c bzw. b und c in unterschiedlicher Anzahl vorkommen.

   Da für jedes Wort in L aber entweder $\#_a(w) \neq \#_b(w)$ oder $\#_b(w) \neq \#_c(w)$ gilt (denn sonst wäre $\#_a(w) = \#_b(w) = \#_c(w)$), reicht es aus, diese zwei Fälle zu unterscheiden. Eine solche vereinfachte Grammatik ist z.B. $G' = (V, \Sigma, P', S\}$ mit $V = \{S, T_{ab}, T_{bc}, A, B, C\}$ und den Produktionen P':

$$\begin{aligned} S &\to T_{ab}C \mid AT_{bc} \\ T_{ab} &\to aT_{ab}b \mid aA \mid bB \\ T_{bc} &\to bT_{bc}c \mid bB \mid cC \\ A &\to aA \mid \epsilon \\ B &\to bB \mid \epsilon \\ C &\to cC \mid \epsilon \end{aligned}$$

2. Ableitungen für G':

   abbcc: $S \to T_{ab}C \to T_{ab}cC \to T_{ab}ccC \to T_{ab}cc \to aT_{ab}bcc \to abBbcc \to abbcc$

   abccc: $S \to AT_{bc} \to aAT_{bc} \to aT_{bc} \to abT_{bc}c \to abcCc \to abccCc \to abccc$

Hausaufgabe 3 (5 Punkte)

Sei $\Sigma = \{0,1\}$. Die zwei Operationen Spiegelung (w^R) und Negation ($\overline{w}$) sind für $w \in \Sigma^*$ wie folgt definiert:

$$w^R = \begin{cases} \epsilon, & \text{falls } w = \epsilon \\ u^R a, & \text{falls } w = au \text{ für } a \in \Sigma \text{ und } u \in \Sigma^* \end{cases}$$

$$\overline{w} = \begin{cases} \epsilon, & \text{falls } w = \epsilon \\ \hat{a}\overline{u}, & \text{falls } w = au \text{ für } a \in \Sigma \text{ und } u \in \Sigma^* \end{cases}$$

Dabei setzen wir $\hat{0} = 1$ und $\hat{1} = 0$. Wie man leicht (etwa per Induktion) zeigen kann, gelten für diese Operationen auch die Gleichungen $(ua)^R = au^R$ und $\overline{ua} = \overline{u}\hat{a}$ für alle $a \in \Sigma$, $u \in \Sigma^*$. Im Folgenden nehmen wir diese Identitäten als bewiesen an.
Wir betrachten nun die Sprache $L = \{w \in \Sigma^* \mid w^R = \overline{w}\}$ und die Grammatik

$$G = (\{S\}, \Sigma, \{S \to 0S1 \mid 1S0 \mid \epsilon\}, S).$$

Zeigen Sie: L ist genau die von der Grammatik G beschriebene Sprache.

Lösungsvorschlag

Wie in der Vorlesung gezeigt, gehört zu der Grammatik G auch eine induktive Definition von $L(G)$ als kleinste Menge, die die folgenden Eigenschaften aufweist:

$$\epsilon \in L(G) \qquad w \in L(G) \Rightarrow 0w1 \in L(G) \qquad w \in L(G) \Rightarrow 1w0 \in L(G)$$

Daraus ergibt sich auch das folgende Induktionsprinzip über die Ableitung eines Wortes w in G: Sei Q eine Eigenschaft von Wörtern aus Σ^*. Gelten $Q(\epsilon)$ und für alle Wörter $w \in \Sigma^*$ sowohl „$Q(w) \Rightarrow Q(0w1)$“ als auch „$Q(w) \Rightarrow Q(1w0)$“, dann gilt $Q(w)$ für alle $w \in L(G)$.

Wir müssen $L = L(G)$ zeigen. Dazu zeigen wir die Aussagen „$w \in L(G) \Rightarrow w \in L$“ und „$w \in L \Rightarrow w \in L(G)$“ für alle $w \in \Sigma^*$.

1. $w \in L(G) \Rightarrow w \in L$. Wir beweisen diesen Fall per Induktion über die Ableitung von w in G. Dabei ist $Q(w)$ die Eigenschaft „$w \in L$“.

   - $\underline{w = \epsilon}$: Es gilt $\epsilon^R = \epsilon = \overline{\epsilon}$ und damit $\epsilon \in L$.
   - $\underline{w \in L \Rightarrow 0w1 \in L}$: Es gilt

     $$(0w1)^R = (w1)^R 0 = 1w^R 0 \overset{w \in L}{=} 1\overline{w}0 = \hat{0}\overline{w}\hat{1} = \overline{0w}\hat{1} = \overline{0w1}$$

     und damit $0w1 \in L$.
   - $\underline{w \in L \Rightarrow 1w0 \in L}$: Dieser Fall wird vollkommen analog zum vorherigen Fall bewiesen, wobei 0 und 1 vertauscht werden.

2. $w \in L \Rightarrow w \in L(G)$. Beweis per Induktion über die Länge von w.

   - $\underline{|w| = 0}$: Dann ist $w = \epsilon$ und damit $w \in L(G)$.

- $|w| = 1$: Dann ist $w = w^R$, aber $w \neq \overline{w}$ (denn $\hat{0} \neq 0$ und $\hat{1} \neq 1$). Also ist $w \notin L$, und damit gilt auch „$w \in L \Rightarrow w \in L(G)$“.
- $|w| \geq 2$: Dann hat w die Form $w = aub$ für gewisse $a, b \in \Sigma$ und $u \in \Sigma^*$, und es gilt
$$bu^Ra = (ub)^Ra = (aub)^R \overset{w\in L}{=} \overline{aub} = \hat{a}\overline{ub} = \hat{a}\overline{u}\hat{b}$$
und damit $b = \hat{a}$ und $u^R = \overline{u}$. Da $|u| < |w|$ folgt $u \in L(G)$ nach Induktionshypothese. Weiterhin ist $w = aub = au\hat{a}$, also entweder $w = 0u1$ oder $w = 1u0$, und mit $u \in L(G)$ gilt dann $w \in L(G)$.

Hausaufgabe 4 (5 Punkte)

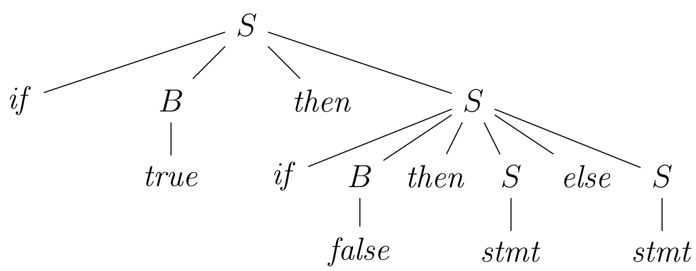
Die Syntax imperativer Programmiersprachen kann in vielen Fällen durch eine kontextfreie Grammatik beschrieben werden. Wir betrachten hier eine eingeschränkte Sprache, die nur aus zwei *if*-Varianten und einer (unspezifizierten) Anweisung besteht.

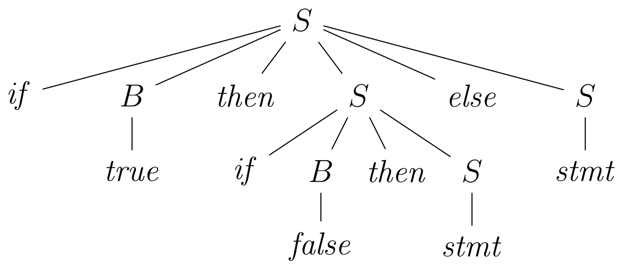
1. Zeigen Sie, dass die Grammatik G mit den folgenden Produktionen nicht eindeutig ist, indem Sie für ein Wort mindestens zwei unterschiedliche Syntaxbäume in G angeben:
$$\begin{aligned} S &\to \mathit{stmt} \mid \mathit{if}\ B\ \mathit{then}\ S \mid \mathit{if}\ B\ \mathit{then}\ S\ \mathit{else}\ S \\ B &\to \mathit{true} \mid \mathit{false} \end{aligned}$$

2. Wir betrachten nun die Grammatik G' mit den folgenden Produktionen:
$$\begin{aligned} S &\to \mathit{if}\ B\ \mathit{then}\ S \mid T \\ T &\to \mathit{if}\ B\ \mathit{then}\ T\ \mathit{else}\ S \mid \mathit{stmt} \\ B &\to \mathit{true} \mid \mathit{false} \end{aligned}$$
Entscheiden Sie, ob diese Grammatik ein- oder mehrdeutig ist und begründen Sie.

Lösungsvorschlag

1. Das folgende Programm kann auf zwei Arten in G abgeleitet werden:
$$\mathit{if\ true\ then\ if\ false\ then\ stmt\ else\ stmt}$$

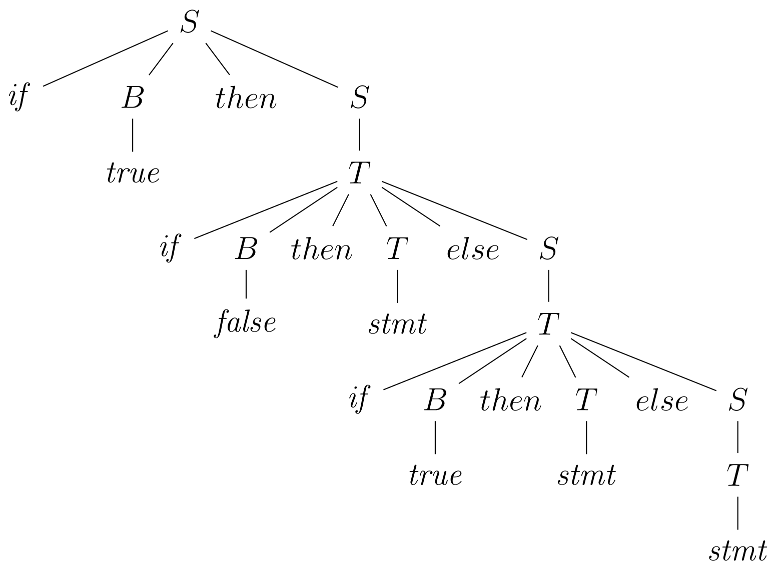


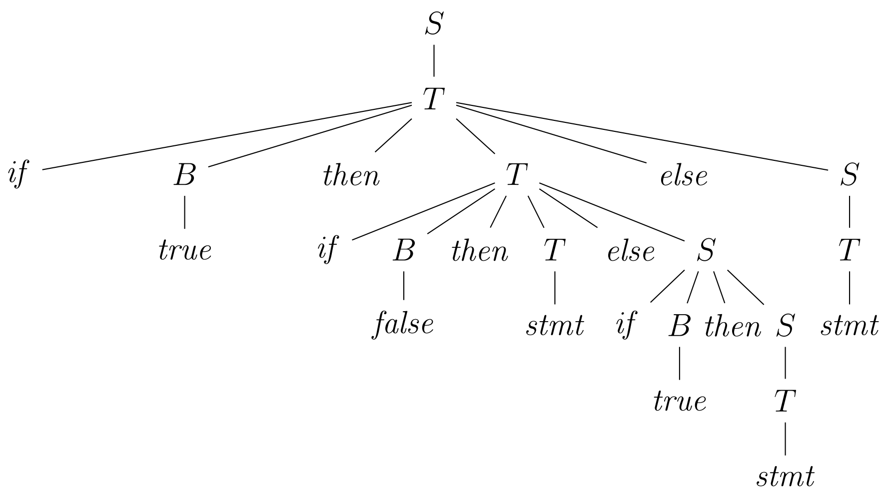


2. Auch diese Grammatik ist nicht eindeutig, denn das Wort

   if true then if false then stmt else if true then stmt else stmt

   kann auf zwei verschiedene Arten in G' abgeleitet werden:





Quiz 1

Beantworten Sie kurz die folgenden Fragen:

1. Kann man mit einem regulären Ausdruck kontextfreie Grammatiken beschreiben?
2. Wie kann man entscheiden, ob die Sprache einer CFG endlich ist?

Lösungsvorschlag

1. Ja. Zum Beispiel mit $\alpha = (S \mid A \mid B) \rightarrow ((S \mid A \mid B \mid a \mid b)^+ \mid \hat{\epsilon})$ kann man eine Produktion darstellen, wobei S, A und B die zulässigen Nichtterminale und a und b die zulässigen Eingabezeichen sind und $\hat{\epsilon}$ für eine leere rechte Seiter einer Produktion steht. Mit einem Trennzeichen, zum Beispiel einem Komma, kann man eine Liste von Produktionen beschreiben, also durch $\alpha(, \alpha)^*$.
2. Analog zu endlichen Automaten überprüft man, ob erreichbare, produktive Nichtterminale an Zyklen beteiligt sind. Die von der Grammatik beschriebene Sprache ist genau dann endlich, wenn dies nicht der Fall ist.

Tutoraufgabe 1

Sei $\Sigma = \{a, b\}$.

1. Finden Sie eine Grammatik G, so dass $L(G) = \{w \in \Sigma^* \mid \#_a(w) = 2 \cdot \#_b(w)\}$.
2. Zeigen Sie die Korrektheit Ihrer Grammatik, d.h., zeigen Sie, dass für alle ableitbaren Wörter $w \in L(G)$ die Beziehung $\#_a(w) = 2 \cdot \#_b(w)$ gilt.
3. Zeigen Sie, dass alle Wörter $(ab)^n a^n$ für $n \geq 0$ in G ableitbar sind.

Lösungsvorschlag

1. Eine mögliche Lösung ist die Grammatik $G = (\{S\}, \Sigma, P, S)$ mit den folgenden Produktionen:
$$S \rightarrow aSaSb \mid aSbSa \mid bSaSa \mid SS \mid \epsilon$$
2. Induktion über die Erzeugung von $w \in L(S)$:
   - $\underline{S \rightarrow \epsilon}$: Dann ist $w = \epsilon$, und die Behauptung gilt offensichtlich.
   - $\underline{S \rightarrow aSaSb}$: Sei $w = auavb$ mit $u, v \in L(S)$. Nach Induktionshypothese gilt die Behauptung für u und v, und damit ist auch
   $$\#_a(w) = \#_a(u) + \#_a(v) + 2 = 2 \cdot (\#_b(u) + \#_b(v) + 1) = 2 \cdot \#_b(w)$$
   - Die Fälle $S \rightarrow aSbSa$ und $S \rightarrow bSaSa$ sowie $S \rightarrow SS$ sind analog.
3. Induktion über n:
   - $\underline{n = 0}$: Dann gilt $(ab)^0 a^0 = \epsilon$, und mit der Produktion $S \rightarrow \epsilon$ ist das leere Wort ableitbar in G.
   - $\underline{n \rightarrow n+1}$: Nach Induktionshypothese gilt $S \rightarrow_G^* (ab)^n a^n$. Dann gilt auch
   $$S \rightarrow_G aSbSa \rightarrow_G abSa \rightarrow_G^* ab(ab)^n a^n a \rightarrow_G (ab)^{n+1} a^{n+1}$$

Tutoraufgabe 2

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

1. Geben Sie eine kontextfreie Grammatik an, die die Menge der regulären Ausdrücke über dem Alphabet $\{0, 1\}$ beschreibt.

2. Ist Ihre Grammatik eindeutig? Falls nicht, geben Sie eine eindeutige Grammatik an, die die Bindungsstärken in regulären Ausdrücken respektiert (also Konkatenation stärker als + bindet).

3. Geben Sie den Syntaxbaum für das Wort 01^*0+1 mit Ihrer eindeutigen Grammatik an.

Lösungsvorschlag

1. 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:

$$R \to 0 \mid 1 \mid \boldsymbol{\epsilon} \mid \boldsymbol{\emptyset} \mid RR \mid R + R \mid R^* \mid (R)$$

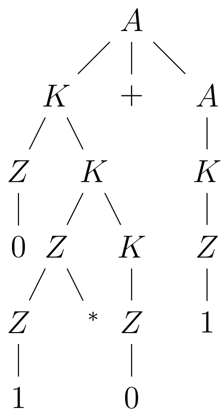
2. Die obige Grammatik ist nicht eindeutig, da z.B. der Ausdruck $0 + 00$ zwei Syntaxbäume hat, die jeweils der Klammerung $0 + (00)$ bzw. $(0 + 0)0$ entsprechen. Eine eindeutige Grammatik erhält man, indem man mehrere Nichtterminale verwendet ähnlich wie in Beispiel 3.3 aus der Vorlesung.

   Wir setzen $G' = (\{A, K, Z\}, \Sigma, P', A)$ mit den Produktionen

$$\begin{aligned} A &\to K \mid K + A \\ K &\to Z \mid ZK \\ Z &\to 0 \mid 1 \mid \boldsymbol{\epsilon} \mid \boldsymbol{\emptyset} \mid (A) \mid Z^* \end{aligned}$$

   Diese Grammatik ist eindeutig, denn jedes Wort w in $L(G')$ ist entweder eine Alternative oder es ist keine Alternative. Es steht also fest, ob die erste angewandte Produktion $A \to K + A$ oder $A \to K$ ist, und nur eine der beiden Produktionen kann als erstes angewandt werden. Entsprechendes gilt für $w \in L(K)$ und die Konkatenation. Bei allen Ableitungen von Wörtern $w \in L(Z)$ aus der Variablen Z ist ebenfalls die erste Produktionsanwendung durch w vorgegeben.

3. Der Syntaxbaum sieht mit G' wie folgt aus:



Tutoraufgabe 3

Sei $G = (V, \Sigma, P, S)$ eine kontextfreie Grammatik.
Zeigen Sie, dass für alle $\alpha, \beta, u, v \in (\Sigma \cup V)^*$ gilt:

1. $\alpha \rightarrow_G \beta \implies u\alpha v \rightarrow_G u\beta v$ (Abschluss von $\rightarrow_G$ unter Kontext)

2. $\alpha \rightarrow_G^n \beta \implies u\alpha v \rightarrow_G^n u\beta v$ (Abschluss von $\rightarrow_G^n$ unter Kontext)

Lösungsvorschlag

1. Wenn $\alpha \rightarrow_G \beta$, dann gibt es nach Def. 3.4 eine Regel $A \rightarrow \gamma \in P$ und Wörter α_1, α_2 (über $V \cup \Sigma$), so dass $\alpha = \alpha_1 A \alpha_2$ und $\beta = \alpha_1 \gamma \alpha_2$. Dann gilt aber auch $u\alpha v = (u\alpha_1)A(\alpha_2 v)$ und $u\beta v = (u\alpha_1)\gamma(\alpha_2 v)$ und damit $u\alpha v \rightarrow_G u\beta v$.

2. Die zuvor gezeigte Eigenschaft übertragen wir mit Induktion über n auf die Mehrschrittrelation $\rightarrow_G^n$:

   $\underline{n = 0}$: Aus $\alpha \rightarrow_G^0 \beta$ folgt $\alpha = \beta$, also $u\alpha v = u\beta v$ und daher $u\alpha v \rightarrow_G^0 u\beta v$.

   $\underline{n \rightarrow n+1}$: Wenn $\alpha \rightarrow_G^{n+1} \beta$, dann $\alpha \rightarrow_G^n \gamma \rightarrow_G \beta$ laut Definition von $\rightarrow_G^n$. Allgemeiner gilt dann auch $u\alpha v \rightarrow_G^n u\gamma v \rightarrow_G u\beta v$ aufgrund der Induktionshypothese und des Abschlusses von $\rightarrow_G$ unter Kontext. Damit folgt dann $u\alpha v \rightarrow_G^{n+1} u\beta v$ aus der Definition von $\rightarrow_G^n$.