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

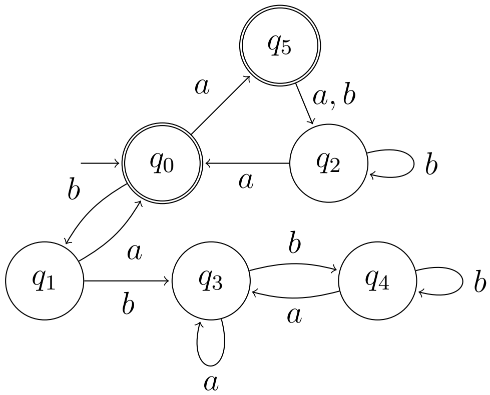
Sommersemester 2011
Lösungsblatt 6
27. Juni 2011

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 (4 Punkte)

Konstruieren Sie durch schrittweises Lösen entsprechender Gleichungen einen regulären Ausdruck für die Sprache $L(M)$ des folgenden DFA $M = (Q, \{a, b\}, \delta, q_0, \{q_0, q_5\})$ mit $Q = \{q_0, q_1, q_2, q_3, q_4, q_5\}$:



Verwenden Sie beim Lösen nur jeweils einen elementaren Schritt und machen Sie deutlich, aus welchen Gleichungen neue Gleichungen abgeleitet wurden.

Lösungsvorschlag

Die entsprechenden Gleichungen sind:

$$X_0 \equiv aX_5 \mid bX_1 \mid \boldsymbol{\epsilon} \tag{1}$$
$$X_1 \equiv aX_0 \mid bX_3 \tag{2}$$
$$X_2 \equiv aX_0 \mid bX_2 \tag{3}$$
$$X_3 \equiv aX_3 \mid bX_4 \tag{4}$$
$$X_4 \equiv aX_3 \mid bX_4 \tag{5}$$
$$X_5 \equiv (a \mid b)X_2 \mid \boldsymbol{\epsilon} \tag{6}$$

Anwendung von Ardens Lemma auf (5) ergibt:

$$X_4 \equiv b^*aX_3$$

Damit folgt für (4):

$$X_3 \equiv aX_3 \mid bb^*aX_3 \equiv (a \mid bb^*a)X_3 \equiv (a \mid bb^*a)X_3 \mid \boldsymbol{\emptyset} \overset{\text{Arden}}{\equiv} (a \mid bb^*a)^*\boldsymbol{\emptyset} \equiv \boldsymbol{\emptyset}$$

Einsetzen in (2) liefert dann:

$$X_1 \equiv aX_0$$

und daher ergibt sich mit (1):

$$X_0 \equiv aX_5 \mid baX_0 \mid \boldsymbol{\epsilon} \tag{7}$$

Aus (3) folgt mit Ardens Lemma:

$$X_2 \equiv b^*aX_0$$

und in Kombination mit (6) ergibt sich daraus:

$$X_5 \equiv (a \mid b)b^*aX_0 \mid \boldsymbol{\epsilon}$$

Einsetzen in (7) liefert dann:

$$\begin{aligned} X_0 &\equiv a((a \mid b)b^*aX_0 \mid \boldsymbol{\epsilon}) \mid baX_0 \mid \boldsymbol{\epsilon} \\ &\equiv (a(a \mid b)b^*a \mid ba)X_0 \mid a \mid \boldsymbol{\epsilon} \\ &\equiv (a(a \mid b)b^*a \mid ba)^*(a \mid \boldsymbol{\epsilon}) && \text{(mit Ardens Lemma)} \end{aligned}$$

Hausaufgabe 2 (5 Punkte)

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

$$\begin{aligned} S &\to TS \mid U \\ T &\to aSb \mid TT \mid c \\ U &\to cT \mid a \end{aligned}$$

Konstruieren Sie eine zu G äquivalente Grammatik G' in Chomsky-Normalform. Bestimmen Sie mit dem CYK-Algorithmus, ob $ccaab \in L(G')$.

Lösungsvorschlag

Wir überführen G schrittweise in Chomsky-Normalform:

1. Elimination von ϵ-Produktionen: Es gibt keine ϵ-Produktionen, die eliminiert werden müssten.

2. Elimination von Kettenregeln: Die Kettenproduktion $S \to U$ muss eliminiert werden:
$$S \to TS \mid cT \mid a \qquad T \to aSb \mid TT \mid c$$

3. Einführung von Nichtterminalen für Buchstaben: Für a, b und c werden Nichtterminale A, B und C eingeführt, da die rechte Seiten der Produktionen $S \to cT$ und $T \to aSb$ eine Länge ≥ 2 haben:

$$S \to TS \mid CT \mid a \qquad T \to ASB \mid TT \mid c \qquad A \to a \qquad B \to b \qquad C \to c$$

4. Aufspalten langer Produktionen:

$$\begin{aligned} S &\to TS \mid CT \mid a & \qquad A &\to a \\ T &\to AT' \mid TT \mid c & B &\to b \\ T' &\to SB & C &\to c \end{aligned}$$

Das Ergebnis ist in Chomsky-Normalform. Nach dem CYK-Algorithmus ergibt sich die folgende Berechnungstabelle:

15 S, T				
14 ∅	25 S, T			
13 S	24 ∅	35 T		
12 S, T	23 S	34 ∅	45 T'	
11 C, T	22 C, T	33 A, S	44 A, S	55 B
c	c	a	a	b

Also ist $ccaab \in L(G')$.

Hausaufgabe 3 (5 Punkte)

Zeigen Sie mit Hilfe des Pumping-Lemmas, dass die Sprache $L = \{a^i b^{(2^i)} \mid i \in \mathbb{N}\}$ nicht kontextfrei ist.

Lösungsvorschlag

Angenommen, L wäre kontextfrei. Dann sei n eine Pumping-Lemma-Zahl und betrachte das Wort $z = a^n b^{(2^n)}$. Es existiert eine Zerlegung $z = uvwxy$, so dass $vx \neq \epsilon$, $|vwx| \leq n$ und für alle $i \in \mathbb{N}$ gilt $uv^i wx^i y \in L$.
Wir unterscheiden vier Fälle:

- $v = a^k$ *und* $x = a^l$ für $k + l > 0$. Dann ist $uv^0wx^0y = a^{n-(k+l)}b^{(2^n)} \notin L$.
- $v = b^k$ *und* $x = b^l$ für $k + l > 0$. Dann ist $uv^0wx^0y = a^n b^{(2^n-(k+l))} \notin L$.
- $v = a^k b^l$ *oder* $x = a^k b^l$ für $k, l > 0$. Dann ist $v^2 = a^k b^l a^k b^l$, enthält also einen b-a-Übergang. Also ist $uv^2wx^2y \notin L$.
- Sonst müssen $v = a^k$ und $x = b^l$ sein für $k, l > 0$. Also ist $uv^{i+1}wx^{i+1}y = a^{n+ik}b^{2^n+il}$ für ein beliebiges $i \in \mathbb{N}$. Dieses Wort ist genau dann in L, wenn gilt $2^{n+ik} = 2^n + il$. Eine lineare und eine exponentielle Funktion haben aber höchstens zwei Schnittpunkte; also gilt obige Gleichung höchstens für zwei i und damit insbesondere nicht für alle i.

Für das Wort $z \in L$ existiert also keine pumpbare Zerlegung, was ein Widerspruch zur Annahme ist. Also ist L nicht kontextfrei.

Hausaufgabe 4 (3 Punkte)

Sei $G = (V, \Sigma, P, S)$ eine kontextfreie Grammatik. Finden Sie einen Algorithmus um die Menge $E = \{N \in V \mid N \to^* \epsilon\}$ zu berechnen und begründen Sie seine Korrektheit.

Lösungsvorschlag

1. Setze $i := 1$ und $E_1 := \{N \in V \mid N \to \epsilon \in P\}$.
2. Solange $i < |V|$
   (a) Setze $i := i + 1$
   (b) Setze $E_i := E_{i-1} \cup \{N \in V \mid N \to \alpha \in P \text{ für ein } \alpha \in E_{i-1}^*\}$.
3. Gebe E_i zurück.

Der Algorithmus terminiert, denn V ist endlich und damit werden genau $|V|$ Schleifendurchläufe ausgeführt.
Einen Syntaxbaum, der das Wort ϵ beschreibt, nennen wir ϵ-Baum. Die Tiefe eines Baumes ist die Anzahl der Kanten auf dem längsten Pfad von der Wurzel bis zu einem Blatt. Für die Korrektheit überlegen wir uns, das für jedes Nichtterminal $N \in E$ ein „flachster" ϵ-Baum existieren muss, der eine Tiefe von höchstens $|V|$ hat: Denn ansonsten gibt es einen Pfad von der Wurzel zu einem Blatt, auf dem ein Nichtterminal zwei mal vorkommt. Dann kann man aber den Pfad verkürzen, indem die beiden Vorkommen dieses Nichtterminals zusammenfasst und alles dazwischen entfernt.
Wir überzeugen uns jetzt per Induktion, dass E_i vor und nach jedem Schleifendurchlauf die folgende Bedingung (†) erfüllt: „E_i enthält die Nichtterminale, zu denen ein ϵ-Baum der Tiefe höchstens i existiert." Mit unseren Vorüberlegungen ist dann $E = E_{|V|}$ und damit folgt die Korrektheit des Algorithmus.
Vor der ersten Iteration ist $i = 1$ und E_1 enthält genau die Nichtterminale, aus denen ϵ direkt, also in einem Schritt abgeleitet werden kann. Daher gilt (†) für $i = 1$.
Für den Schleifendurchlauf überlegen wir uns folgendes: Für ein $i > 1$ existiert ein ϵ-Baum der Tiefe i genau dann, wenn ein Wort $\alpha \in V^*$ existiert, so dass zu jedem Nichtterminal in α ein ϵ-Baum der Tiefe $i - 1$ existiert.
Da wir (nach Induktion) annehmen können, dass (†) für E_{i-1} gilt, folgt also: $N \to^i \epsilon$ genau dann, wenn $N \to \alpha \in P$ und $\alpha \in E_{i-1}^*$.

Hausaufgabe 5 (3 Punkte)

Wir betrachten Grammatiken, bei denen jede Produktion die Form $A \to aB$ oder $A \to Ba$ oder $A \to \epsilon$ für ein Terminal a und Nichtterminale A, B hat. Diese Grammatiken beschreiben die L-Sprachen. Offensichtlich gilt „regulär $\subseteq$ L-Sprachen $\subseteq$ kontextfrei".

1. Entscheiden Sie: Gilt entweder „regulär $\subset$ L-Sprachen" oder „regulär $=$ L-Sprachen"? Beweisen Sie!
2. Entscheiden und begründen Sie: Gilt „L-Sprachen $=$ kontextfrei"?

Lösungsvorschlag

Die L-Sprachen heißen auch „lineare Sprachen". Es gilt „regulär $\subset$ linear $\subset$ kontextfrei".

1. *regulär $\subset$ linear*: Die Sprache $\{a^n b^n \mid n \in \mathbb{N}\}$ ist bekanntermaßen nicht regulär. Sie ist aber linear, denn sie wird von der linearen Grammatik $(\{A, B\}, \{a, b\}, P, A)$ mit den Produktionen $A \to aB \mid \epsilon$ und $B \to Ab$ erzeugt.

2. *linear* $\subset$ *kontextfrei*: Die Sprache $\{a^m b^m a^n b^n \mid m, n \in \mathbb{N}\}$ ist kontextfrei, denn sie wird von der kontextfreien Grammatik $(\{S, A\}, \{a, b\}, P, S)$ mit den Produktionen $S \to AA$ und $A \to aAb \mid \epsilon$ erkannt.

   Aber diese Sprache ist nicht linear: Intuitiv liegt das daran, dass bei einer Ableitung einer linearen Grammatik in jedem Schritt genau ein Nichtterminal produziert wird. Wenn ein Fragment $a^k b^l$ mit $k = l$ erzeugt werden soll, liegt das Nichtterminal zwangsläufig zwischen a und b. Daher ist es nicht möglich, die Bedingung $k = l$ für zwei nebeneinanderstehende Paare von as und bs zu erzwingen.

   Formal lässt sich das mit einem speziellen Pumping-Lemma für lineare Sprachen beweisen.

Quiz 1

Beantworten Sie kurz die folgenden Fragen:

1. Ist das Wortproblem für ein Wort der Länge n und eine kontextfreie Grammatik in der Zeit $O(n^4)$ entscheidbar?
2. Sei $\Sigma = \{a, b, c\}$. Wieviele Nichtterminale muss eine CFG in Chomsky-Normalform mindestens haben, so dass sie genau alle Wörter der Länge 3 erzeugt?
3. Sei L kontextfrei und Fragment(L) die Sprache aller Fragmente der Wörter aus L, wobei ein Fragment durch Weglassen beliebiger Teile eines Wortes entsteht. Ist diese Sprache kontextfrei?

Lösungsvorschlag

1. Laut Vorlesung entscheidet der CYK-Algorithmus das Wortproblem in $O(n^3)$ und damit insbesondere auch in $O(n^4)$.
2. Mindestens drei Nichtterminale, wie zum Beispiel die Grammatik mit den folgenden Produktionen:
$$S \to TU \qquad T \to UU \qquad U \to a \mid b \mid c$$
Mit nur einem Nichtterminal weniger werden auch Wörter kürzerer Länge erzeugt.
3. Fragment(L) wurde für reguläre Sprachen bereits in Übung 2 betrachtet. Das Vorgehen hier ist ähnlich: Ist G eine Grammatik für L mit Produktionen P, so erhält man eine Grammatik G' für Fragment(L), indem man P durch
$$P' = \{N \to \beta \mid N \to \alpha \in P \text{ und } \beta \text{ Fragment von } \alpha\}$$
ersetzt.

Tutoraufgabe 1

Zeigen Sie, dass $L = \{a^n b^m c^k \mid 0 < n < m < k\}$ keine kontextfreie Sprache ist.

Lösungsvorschlag

Wir nehmen an, dass L kontextfrei ist und führen diese Annahme zum Widerspruch.
Sei $n \geq 1$ eine Pumping-Lemma-Zahl für die kontextfreie Sprache L. Sei zusätzlich $z = a^n b^{n+1} c^{n+2}$, d.h., $z \in L$ und $|z| \geq n$. Gemäß Pumping-Lemma gibt es dann also eine Zerlegung $z = uvwxy$ mit Wörtern $u, v, w, x, y \in \{a, b, c\}^*$ und den folgenden Eigenschaften:

$$(1)\ vx \neq \epsilon \qquad (2)\ |vwx| \leq n \qquad (3)\ \forall i \in \mathbb{N}.\ uv^i wx^i y \in L$$

Aufgrund von (1) unterscheiden wir die folgenden Fälle:

- $\#_a(vx) > 0$: Wegen (2) gilt $\#_c(vx) = 0$. Allerdings gilt auch:
$$\#_a\left(uv^3wx^3y\right) = n + 2\#_a(vx) \geq n + 2 = \#_c\left(uv^3wx^3y\right)$$
Dies ist ein Widerspruch zu (3).

- $\#_a(vx) = 0$ und $\#_b(vx) > 0$: Dann gilt:

$$\#_b\left(uv^0wx^0y\right) = n + 1 - \#_b(vx) < n = \#_a\left(uv^0wx^0y\right)$$

  Daher ist $uv^0wx^0y \notin L$, ein Widerspruch zu (3).

- $\#_a(vx) = 0$ und $\#_b(vx) = 0$: Dann muss $\#_c(vx) > 0$ gelten, und es folgt:

$$\#_c\left(uv^0wx^0y\right) = n + 2 - \#_c(vx) < n + 1 = \#_b\left(uv^0wx^0y\right)$$

  Daher ist $uv^0wx^0y \notin L$, ein Widerspruch zu (3).

Da jeder Fall zu einem Widerspruch führt und die obigen Fälle alle möglichen Zerlegungen abdecken, kann die ursprüngliche Annahme nicht gelten. Also ist L nicht kontextfrei.

Tutoraufgabe 2

Konstruieren Sie eine Turingmaschine $T = (Q, \Sigma, \Gamma, \delta, q_0, \Box, F)$, mit $\Sigma = \{|\}$, die eine eingegebene Strichzahl verdoppelt.

Lösungsvorschlag

Zu Beginn steht, außer Leerzeichen, nur eine Sequenz von Strichen auf dem Band. Der Schreib-/Lesekopf der Turing-Maschine steht auf dem ersten Strich (von links gesehen). Die Berechnung erfolgt, indem jeweils der erste Strich (von links gesehen) durch ein Hilfszeichen X ersetzt wird und zusätzlich ein Hilfszeichen X an das linke Ende geschrieben wird. Zum Schluß werden alle Hilfszeichen von rechts nach links durch Striche ersetzt.
Sei $T = (Q, \Sigma, \Gamma, \delta, q_0, \Box, F)$ mit $Q = \{q_0, q_1, q_2, q_3\}$, $\Sigma = \{|\}$, $\Gamma = \{|, X, \Box\}$ und $F = \{q_3\}$.
Die Übergangsfunktion δ kann man der folgenden Tabelle entnehmen:

Übergang	Kommentar
$\delta(q_0, \|) = (q_1, X, L)$	ersetze erstes | durch Hilfszeichen X
$\delta(q_1, X) = (q_1, X, L)$	gehe nach links zum ersten X
$\delta(q_1, \Box) = (q_0, X, R)$	füge zusätzliches Hilfszeichen X am Anfang ein
$\delta(q_0, X) = (q_0, X, R)$	gehe nach rechts zum ersten |
$\delta(q_0, \Box) = (q_2, \Box, L)$	alle | abgearbeitet, Band enthält doppelte Anzahl von X
$\delta(q_2, X) = (q_2, \|, L)$	ersetze X durch |
$\delta(q_2, \Box) = (q_3, \Box, R)$	alle X durch | ersetzt, Stopp