SS 2015

Zentralübung zur Vorlesung Theoretische Informatik

Dr. Werner Meixner

Fakultät für Informatik
TU München

http://www14.in.tum.de/lehre/2015SS/theo/uebung/

23. April 2015

ZÜ II

Übersicht:

1. Übungsbetrieb	Fragen, Probleme
2. Themen	Klammerstrukturen und das Ballot Problem BNF für kontextfreie Sprachen
3. Vorbereitung	TA Blatt 2

1. Fragen, Probleme, Anregungen

Aktuelle Fragen?

Hinweis: Fehler Folie 29, ZÜ 1 siehe Webseite.

2. Themen

2.1 Klammerstrukturen

Um Klammerausdrücke zu studieren, kann man sich zunächst auf Ausdrücke ohne Variable und einen einzigen Klammertyp von öffnenden und schließenden Klammern ,,(" bzw. ,,) " beschränken.

Die Menge $K \subseteq \Sigma^*$ von Wörtern über der Zeichenmenge $\Sigma = \{(,)\}$ heißt Menge der korrekten Klammerausdrücke über Σ, wenn K ausschließlich mit folgenden Regeln erzeugt werden kann.

- $\varepsilon \in K$,
- $w \in K \implies (w) \in K$,
- $w_1, w_2 \in K \implies w_1 w_2 \in K$.

Wir nennen ein Wort $w \in \Sigma^*$ *semipositiv*,
falls $|u|_(\geq |u|_)$, d.h.

$$\Delta(u) := |u|_(- |u|_) \geq 0$$

für alle Anfangsteilwörter u von w gilt.

Dabei bezeichnet $|u|_x$ die Anzahl der Vorkommen des Zeichens x in dem Wort u.

Es gilt Folgendes:

Falls $w \in \Sigma^*$ ein korrekter Klammerausdruck ist,
d.h. nach obigen Regeln gebildet wurde,
dann ist w semipositiv
und es gilt $|w|_(= |w|_)$.

Beweisidee

Jede der drei Regeln wird daraufhin überprüft,
ob die Regelanwendung die Eigenschaft, semipositiv zu sein,
invariant läßt.

Der Beweis für das Umgekehrte ist etwas länger,
siehe aber Grafik bei Ballot-Problem.

Tatsächlich ist die Eigenschaft, semipositiv zu sein,
sogar ein Kriterium für korrekte Klammerbildung.

2.2 Das Ballot-Problem

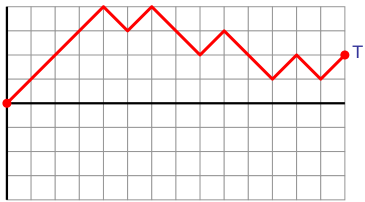
(DS Vorl. WS 12/13 Prof. Mayr)

Bei einer Wahl werden a Stimmen für Kandidat A und b Stimmen für Kandidat B gezählt, mit $a > b \geq 0$. Die Stimmzettel werden sequentiell ausgezählt.

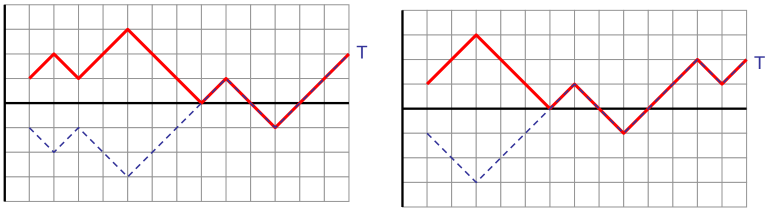
Zählproblem: Wie viele Zählfolgen gibt es, so dass A nach jedem Schritt in Führung ist?

Wir stellen jede Zählfolge durch einen Pfad im ganzzahligen Gitter $\mathbb{N}_0 \times \mathbb{Z}$ dar, der vom Punkt $(0, 0)$ zum Punkt $(a + b, a - b)$ verläuft und bei dem eine Stimme für A (bzw. für B) einer diagonalen Kante um 1 nach rechts und 1 nach oben (bzw. unten) entspricht.

Es ist leicht zu sehen, dass die Anzahl der „guten“ Pfade vom Ursprung zum Punkt $T := (a+b, a-b)$ gleich der Anzahl der „guten“ Pfade vom Punkt $(1,1)$ zum Punkt T ist:



Durch Spiegelung des Anfangsegments eines „schlechten“ Pfades bis zum ersten Zusammentreffen mit der horizontalen Achse an dieser Achse:



erhalten wir, dass die Anzahl der „schlechten“ Pfade von $(1,1)$ zu T gleich der Anzahl aller Pfade von $(1,-1)$ zu T ist.

Damit ergibt sich

Anzahl der „guten" Pfade von $(0,0)$ zu T

$$
\begin{aligned}
&= \text{Anzahl der „guten" Pfade von } (1,1) \text{ zu } T\\
&= \binom{a-1+b}{b} - \binom{(a-1+1)+(b-1)}{a}\\
&= \binom{a+b-1}{b} - \binom{a+b-1}{a}\\
&= \frac{a-b}{a+b}\binom{a+b}{a}\,.
\end{aligned}
$$

Die beiden Binomialkoeffizienten ergeben sich, da wir im ersten Fall $a-1$ Schritte nach rechts oben und b Schritte nach rechts unten haben, im zweiten Fall im Vergleich dazu jedoch a Schritte nach rechts oben und $b-1$ Schritte nach rechts unten haben.

Transformation des Ballot-Problems

Seien $\Sigma = \{(,)\}$ der geordnete Zeichenvorrat mit einer öffnenden bzw. einer schließenden Klammer.

u ist ein *Anfangsteilwort* (Praefix) von w, falls es ein Wort v gibt, so dass $w = uv$ gilt.

Wir nennen ein nicht leeres Wort $w \in \Sigma^*$ *positiv*, falls $|u|_(> |u|_)$ für alle nichtleeren Anfangsteilwörter u von w gilt.

Zählproblem: Wie viele positive Wörter über Σ mit a öffnenden und b schließenden Klammern gibt es, wobei $a > b \geq 0$ gelte!

2.3 BNF für kontextfreie Sprachen

Die Backus-Naur-Form (BNF) ist ein Formalismus zur kompakten Darstellung von kontextfreien Grammatiken.

Wir wiederholen die entsprechenden Folien der Vorlesung.

Die Backus-Naur-Form (BNF) ist ein Formalismus zur kompakten Darstellung von Typ-2-Grammatiken.

- Statt

$$\begin{array}{ccc} A & \rightarrow & \beta_1 \\ A & \rightarrow & \beta_2 \\ & \vdots & \\ A & \rightarrow & \beta_n \end{array}$$

  schreibt man

$$A \rightarrow \beta_1|\beta_2|\ldots|\beta_n \; .$$

Die Backus-Naur-Form (BNF) ist ein Formalismus zur kompakten Darstellung von Typ-2-Grammatiken.

- Statt

$$\begin{array}{rcl} A & \rightarrow & \alpha\gamma \\ A & \rightarrow & \alpha\beta\gamma \end{array}$$

  schreibt man

$$A \rightarrow \alpha[\beta]\gamma\ .$$

(D.h., das Wort β kann, muss aber nicht, zwischen α und γ eingefügt werden.)

Die Backus-Naur-Form (BNF) ist ein Formalismus zur kompakten Darstellung von Typ-2-Grammatiken.

- Statt

$$\begin{array}{lcl} A & \rightarrow & \alpha\gamma \\ A & \rightarrow & \alpha B\gamma \\ B & \rightarrow & \beta \\ B & \rightarrow & \beta B \end{array}$$

  schreibt man

$$A \rightarrow \alpha\{\beta\}\gamma\ .$$

  (D.h., das Wort β kann beliebig oft (auch Null mal) zwischen α und γ eingefügt werden.)

3. Vorbereitung TA Blatt 2

3.1 VA 1

Eine Grammatik G sei gegeben in BNF-Form durch

$$S \to a\,S\,d\,d, \qquad\qquad S \to \{b\} \mid \{c\}.$$

Geben Sie G als kontextfreie Grammatik $G = (V, \{a, b, c, d\}, P, S)$ an.

Lösung

Der Ausdruck $S \rightarrow \{b\} \mid \{c\}$ steht für die beiden Ausdrücke $S \rightarrow \{b\}$ und $S \rightarrow \{c\}$.

Der Ausdruck $S \rightarrow \{b\}$ steht für die Menge von Regeln $\{S \rightarrow \epsilon, S \rightarrow X, X \rightarrow b, X \rightarrow bX\}$.

Der Ausdruck $S \rightarrow \{c\}$ steht für die Menge von Regeln $\{S \rightarrow \epsilon, S \rightarrow Y, Y \rightarrow c, Y \rightarrow cY\}$.

Dabei sind X, Y Variable, die <u>nicht</u> in G vorkommen.

Es bleibt noch die Umbezeichnung von S nach T, um die Monotoniebedingung zu sichern. Deswegen muss auch $T \to add$ hinzugefügt werden.

Wir fassen zusammen.

$$\begin{array}{rcl} V & := & \{S, X, Y, T\}\,, \\ P & := & \{\,T \to aTdd\,,\, T \to add\,, \\ & & \;\; T \to X\,,\, X \to b\,,\, X \to bX\,, \\ & & \;\; T \to Y\,,\, Y \to c\,,\, Y \to cY\,, \\ & & \;\; S \to \epsilon\,,\, S \to T\,\}\,. \end{array}$$

3.2 VA 2

Gegeben sind folgende Grammatiken:

$$G_1 := (\{S\},\ \{\mathtt{a}, \mathtt{b}, \mathtt{+}, \mathtt{(}, \mathtt{)}\},\\ \{S \to \mathtt{a},\ S \to \mathtt{b},\ S \to S\mathtt{+}S,\ S \to \mathtt{(}S\mathtt{)}\},\ S),$$

$$G_2 := (\{S\},\ \{\mathtt{a}, \mathtt{b}, \mathtt{+}, \mathtt{(}, \mathtt{)}\},\\ \{S \to \mathtt{a},\ S \to \mathtt{b},\ S \to \mathtt{a+}S,\ S \to \mathtt{b+}S,\ S \to \mathtt{(}S\mathtt{)}\},\ S).$$

1. Ordnen Sie die Grammatiken in die Chomsky-Hierarchie ein.
2. Geben Sie jeweils einen Ableitungsbaum für das Wort `a+(b+a)` an.
3. Gilt $L(G_1) = L(G_2)$?

1. Ordnen Sie die Grammatiken in die Chomsky-Hierarchie ein.

Lösung

Die Frage der Einordnung einer Grammatik in die Chomsky-Hierarchie ist erst dann vollständig beantwortet, wenn gesagt wird, welchen Typ sie besitzt und gleichzeitig gesagt wird, dass sie nicht den nächsthöheren Typ besitzt.

Wir zeigen, dass beide Grammatiken vom Typ 2 und nicht vom Typ 3 sind.

Damit eine Grammatik vom Typ 2 ist, genügt es nicht, dass alle Regeln die Form $\alpha \to \beta$ mit $\alpha \in V$ haben.

Es muss auch die Monotoniebedingung erfüllt sein, d. h. $|\alpha| \leq |\beta|$ mit der leichten Abschwächung, dass diese Ungleichung nur für $\alpha \neq S$ gefordert wird und

im Fall $\beta = \epsilon$ und $\alpha = S$ gleichzeitig die Bedingung gelten muss, dass S nie auf der rechten Seite einer Produktion vorkommt.

Da für alle Regeln $\alpha \to \beta$ beider Grammatiken $\beta \neq \epsilon$ und außerdem $\alpha \in V$ gilt, folgt trivialerweise sogar die strikte Monotonie der Grammatiken, d. h. $|\alpha| \leq |\beta|$ für alle Regeln.

Deshalb sind wegen $\alpha \in V$ beide Grammatiken vom Typ 2 oder kontextfrei.

Beide Grammatiken sind nicht regulär, d. h. nicht vom Typ 3.

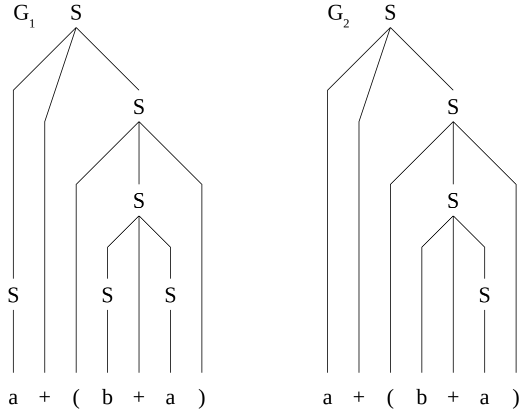
Als Beleg dafür kann die in beiden Grammatiken enthaltene Produktion $S \to (S)$ dienen, die weder rechtslinear noch linkslinear ist.

Allerdings ist G_2 eine lineare Grammatik, G_1 ist nicht linear .

2. Geben Sie jeweils einen Ableitungsbaum für das Wort `a+(b+a)` an.

2. Geben Sie jeweils einen Ableitungsbaum für das Wort `a+(b+a)` an.

Lösung



3. Gilt $L(G_1) = L(G_2)$?

Lösung

Zunächst gilt für $w = \texttt{(a+a)+a}$, dass $w \in L(G_1)$ wegen

$$S \underset{G_1}{\rightarrow} S\texttt{+}S \underset{G_1}{\rightarrow} \texttt{(}S\texttt{)+}S \underset{G_1}{\rightarrow} \texttt{(}S\texttt{+}S\texttt{)+}S \underset{G_1}{\rightarrow}^{*} \texttt{(a+a)+a}\,.$$

Andererseits gilt $w \notin L(G_2)$.
Zum Beweis dieser Aussage nehmen wir $S \underset{G_2}{\longrightarrow}^{*}$ (a+a)+a an.

Im allerersten Ersetzungsschritt kann keine Produktion angewandt werden, die ein Terminalzeichen a oder b am Anfang der hergeleiteten Satzform produziert.
D. h., es kann nur die Produktion $S \rightarrow (S)$ zur Anwendung kommen, also gilt

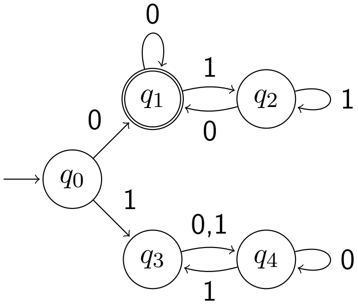
$$S \underset{G_2}{\longrightarrow} (S) \underset{G_2}{\longrightarrow}^{*} \texttt{(a+a)+a}\,.$$

Dies ist aber sicher falsch, denn die schließende Klammer kann in keinem Ersetzungsschritt entfernt werden, mithin gilt
(a+a)+a $\notin L(G_2)$.

Es folgt $L(G_1) \neq L(G_2)$.

3.3 VA 3

Wir betrachten einen endlichen deterministischen Automat $A = (Q, \Sigma, \delta, q_0, F)$, der durch die folgende Grafik gegeben ist.



1. Übersetzen Sie die Grafik in eine extensionale Mengenschreibweise (Darstellung durch Auflistung) für Q, Σ, δ und F.

2. Bestimmen Sie $\delta(\delta(q_1, 0), 1)$ und $\hat{\delta}(q_0, 10)$!

3. Geben Sie ein möglichst einfaches Kriterium an, mit dem man entscheiden kann, ob ein Wort $w \in \Sigma^*$ von A akzeptiert wird.

1. Übersetzen Sie die Grafik in eine extensionale Mengenschreibweise (Darstellung durch Auflistung) für Q, Σ, δ und F.

Lösung

$$
\begin{aligned}
Q &= \{q_0, q_1, q_2, q_3, q_4\}\,, \\
\Sigma &= \{0, 1\}\,, \\
\delta &= \{((q_0,0),q_1),\ ((q_0,1),q_3), \\
&\quad ((q_1,0),q_1),\ ((q_1,1),q_2), \\
&\quad ((q_2,0),q_1),\ ((q_2,1),q_2), \\
&\quad ((q_3,0),q_4),\ ((q_3,1),q_4), \\
&\quad ((q_4,0),q_4),\ ((q_4,1),q_3)\}\,, \\
F &= \{q_1\}\,.
\end{aligned}
$$

Startzustand ist q_0.

2. Bestimmen Sie $\delta(\delta(q_1, 0), 1)$ und $\hat{\delta}(q_0, 10)$!

Lösung

$\delta(\delta(q_1, 0), 1) = \delta(q_1, 1) = q_2$,

$\hat{\delta}(q_0, 10) = \hat{\delta}(\delta(q_0, 1), 0) = \hat{\delta}(q_3, 0) = \hat{\delta}(\delta(q_3, 0), \epsilon) = \delta(q_3, 0) = q_4$.

2. Geben Sie ein möglichst einfaches Kriterium an, mit dem man entscheiden kann, ob ein Wort $w \in \Sigma^*$ von A akzeptiert wird.

Lösung

Wörter, die mit 1 beginnen, werden nicht akzeptiert, weil in dem betreffenden Zweig kein Endzustand erreichbar ist.

Andererseits endet offenbar jedes Wort, das in den Endzustand führt, mit dem Zeichen 0.

Wir vermuten,
dass alle diejenigen Wörter akzeptiert werden, die
mit 0 beginnen und mit 0 enden.

Sei $w = 0$.
Offenbar beginnt w mit 0 und endet gleichzeitig mit 0. Und A akzeptiert w.

Nun sei $w = 0u0$ ein Wort mit $u \in \Sigma^*$.

Die Eingabe von 0 führt in den Zustand q_1.
Von q_1 aus kommt man bei beliebiger Eingabe nur entweder nach q_1 oder q_2.

Gleich wo man gelandet ist, führt die Eingabe von 0 wieder in den Endzustand q_1. D. h., dass w akzeptiert wird.

3.4 VA 4

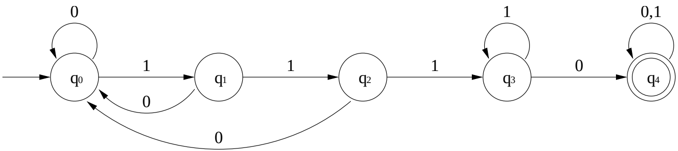
Geben Sie jeweils einen endlichen Automat (als Graph und Übergangsrelation) an, der über dem Alphabet $\Sigma = \{0, 1\}$ folgende Sprache akzeptiert:

1. Die Menge aller Wörter, die das Teilwort 1110 enthalten.
2. Die Menge aller Wörter, bei denen die Anzahl der Einsen durch 3 teilbar ist.
3. Die Menge aller Wörter, die mit 10 beginnen und auf 01 enden.

1. Die Menge aller Wörter, die das Teilwort 1110 enthalten.

Lösung

Die Menge aller Wörter, die das Teilwort 1110 enthalten.

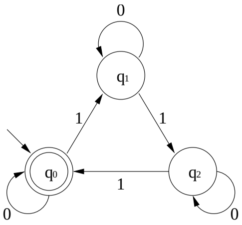


Übergangsrelation:

q_i	$\delta(q_i, 0)$	$\delta(q_i, 1)$
q_0	q_0	q_1
q_1	q_0	q_2
q_2	q_0	q_3
q_3	q_4	q_3
q_4	q_4	q_4

2. Die Menge aller Wörter, bei denen die Anzahl der Einsen durch 3 teilbar ist.

Lösung

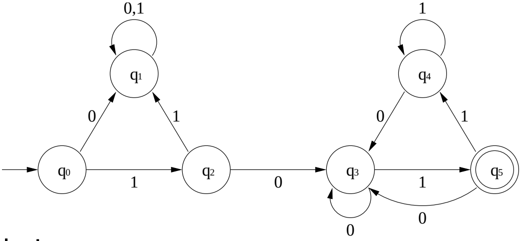


Übergangsrelation:

q_i	$\delta(q_i, 0)$	$\delta(q_i, 1)$
q_0	q_0	q_1
q_1	q_1	q_2
q_2	q_2	q_0

3. Die Menge aller Wörter, die mit 10 beginnen und auf 01 enden.

Lösung



Übergangsrelation:

q_i	$\delta(q_i, 0)$	$\delta(q_i, 1)$
q_0	q_1	q_2
q_1	q_1	q_1
q_2	q_3	q_1
q_3	q_3	q_5
q_4	q_3	q_4
q_5	q_3	q_4