Dr. Frank Drewes
Studiengang Informatik
MZH 3270, Tel.: 4335
E-Mail: drewes@informatik.uni-bremen.de

12. Januar 1998

Theoretische Informatik 2

– Ergänzungen zum Skript (2) –

Linksableitungen und Kellerautomaten

Da die Abschnitte 10 und 11 des Skripts noch recht unvollständig sind und zudem kleine Fehler enthalten, hier eine Ergänzung, die einige der Lücken schließt und die Fehler (hoffentlich) vermeidet. Beachte, daß die Definition von Linksableitungen und das dazugehörige Lemma wie in der Vorlesung angegeben lauten, was eine leichte Abweichung von den im Skript zu findenden Versionen bedeutet. Was den Beweis des Lemmas anbelangt, so sei außerdem darauf hingewiesen, daß er auf der Benutzung des Kontextfreiheitslemmas beruht, welches im Skript erst in Abschnitt 12 behandelt wird, in der Vorlesung aber (aus genau diesem Grunde) vorgezogen wurde.

Definition (Linksableitung)

Sei $G = (N, T, P, S)$ eine kontextfreie Grammatik. Eine *Linksableitung* ist eine Ableitung $u_1 \to u_2 \to \cdots \to u_n$, bei der in jedem Schritt $u_i \to u_{i+1}$ $(1 \leq i < n)$ das am weitesten links stehende Nichtterminal ersetzt wird.

Das folgende Lemma besagt, daß man sich bei kontextfreien Grammatiken auf die Betrachtung von Linksableitungen beschränken kann, ohne daß dies eine Auswirkung auf die erzeugte Sprache hat.

Lemma

Sei $G = (N, T, P, S)$ eine kontextfreie Grammatik. Dann läßt sich jede Ableitung $S \overset{*}{\underset{P}{\to}} v$ mit $v \in T^*$ in eine Linksableitung $S \overset{*}{\underset{P}{\to}} v$ umformen.

Beweis Per Induktion über die Länge von Ableitungen wird die etwas allgemeinere Aussage bewiesen, daß sich jede Ableitung $A \overset{*}{\to} v$ mit $A \in N$, $v \in T^*$ in eine Linksableitung umformen läßt. Für Ableitungen der Länge Null (Induktionsanfang) ist dies trivial, denn eine Ableitung $A \overset{0}{\to} v$ existiert nicht. Betrachte daher jetzt eine Ableitung $A \to u \overset{n}{\to} v$ und nehme an, die Behauptung gelte für Ableitungen der Länge $\leq n$. Sei $u = u_0 A_1 u_1 \cdots A_m u_m$ mit $u_0, \ldots, u_m \in T^*$ und $A_1, \ldots, A_m \in N$. Aufgrund des Kontextfreiheitslemmas gilt $A_i \overset{n_i}{\to} v_i$ für $i = 1, \ldots, m$, wobei $n_i \leq n$ (da $\sum_{i=1}^m n_i = n$) und $u_0 v_1 u_1 \cdots v_m u_m = v$. Nach Induktionsvoraussetzung kann jede der Ableitungen $A_i \overset{n_i}{\to} v_i$ in eine Linksableitung $A_i \overset{n_i}{\to} v_i$ umgeformt werden. Unter

Benutzung dieser Linksableitungen erhält man eine Ableitung

$$\begin{array}{rcl} A & \to & u_0 A_1 u_1 \cdots A_m u_m \\ & \overset{*}{\to} & u_0 v_1 u_1 A_2 u_2 \cdots A_m u_m \\ & \overset{*}{\to} & u_0 v_1 u_1 v_2 u_2 A_3 u_3 \cdots A_m u_m \\ & \vdots & \\ & \overset{*}{\to} & u_0 v_1 u_1 \cdots v_m u_m, \end{array}$$

die nach Definition ebenfalls eine Linksableitung ist. □

Im Skript wird der zu einer kontextfreien Grammatik $G = (N, T, P, S)$ gehörige Kellerautomat $PDA(G)$ definiert als

$$PDA(G) = (\{s_0, OK\}, T, N \cup T \cup \{c_0, c_{OK}\}, d_G, s_0, \{OK\}, c_0)$$

mit

$$\begin{array}{lcl} d_G(s_0, -, c_0) & = & \{(s_0, Sc_{OK})\}, \\ d_G(s_0, -, A) & = & \{(s_0, u) \mid (A ::= u) \in P\}, \\ d_G(s_0, x, x) & = & \{(s_0, \lambda)\} \text{ für alle } x \in T, \text{ und} \\ d_G(s_0, -, c_{OK}) & = & \{(OK, \lambda)\}. \end{array}$$

Im darauffolgenden Korrektheitsbeweis wird $L(G) = L(PDA(G))$ unter Zuhilfenahme folgender Eigenschaften nachgewiesen:

(1) Sei $w = u\overline{u} \in (N \cup T)^*$, wobei u das längste terminale Anfangsstück von w ist. Gibt es eine Linksableitung $S \overset{*}{\to} w$, so gilt $(s_0, uv, Sc_{OK}) \overset{*}{\vdash} (s_0, v, \overline{u}c_{OK})$ für alle $v \in T^*$. (Bemerkung: Im Skript wird irrtümlich nur die Existenz einer beliebigen Ableitung $S \overset{*}{\to} w$ gefordert, während die Behauptung nur für Linksableitungen gilt.)

(2) $(s_0, uv, Sc_{OK}) \overset{*}{\vdash} (s_0, v, \overline{u}c_{OK})$ impliziert $S \overset{*}{\to} u\overline{u}$ für beliebige $u, v \in T^*$ und $\overline{u} \in (N \cup T)^*$.

Die im Skript fehlenden Beweise von (1) und (2) werden hiermit nachgeliefert.

Beweis von (1) (mittels Induktion über die Ableitungslänge)

Für $S \overset{0}{\to} w$ gilt $w = S$, d.h. $u = \lambda$ und $\overline{u} = S$, und daher

$$(s_0, uv, Sc_{OK}) = (s_0, v, \overline{u}c_{OK}) \overset{0}{\vdash} (s_0, v, \overline{u}c_{OK}).$$

Betrachte nun eine Linksableitung $S \overset{n}{\to} u_1 A \overline{u}_1 \to u_1 r \overline{u}_1 = w$ und nehme an, die Behauptung gelte für Ableitungen der Länge n. Da die betrachtete Ableitung eine Linksableitung ist, gilt $u_1 \in T^*$. Also ist u_1 ein terminales Anfangsstück von w und damit auch Anfangsstück von u, d.h. es gibt ein $t \in T^*$ mit $u = u_1 t$ und $t\overline{u} = r\overline{u}_1$. Unter Benutzung der Induktionsvoraussetzung erhält man also die Konfigurationsfolge

$$\begin{array}{rcl} (s_0, uv, Sc_{OK}) & = & (s_0, u_1 tv, Sc_{OK}) \\ & \overset{*}{\vdash} & (s_0, tv, A\overline{u}_1 c_{OK}) \\ & \vdash & (s_0, tv, r\overline{u}_1 c_{OK}) \\ & = & (s_0, tv, t\overline{u}c_{OK}) \\ & \overset{*}{\vdash} & (s_0, v, \overline{u}c_{OK}), \end{array}$$

wobei sich die dritte Zeile aus Zeile 2 der Definition von d_G und die letzte Zeile aus Zeile 3 der Definition von d_G ergibt. $\square$

Beweis von (2) (mittels Induktion über die Länge der Konfigurationsfolgen)

Für $(s_0, uv, Sc_{OK}) \overset{0}{\longmapsto} (s_0, v, \overline{u}c_{OK})$ gilt $u = \lambda$ und $\overline{u} = S$, woraus $S \overset{0}{\to} S = u\overline{u}$ folgt. Nehme jetzt an, die Behauptung gelte für Konfigurationsfolgen der Länge n und betrachte $(s_0, uv, Sc_{OK}) \overset{n+1}{\longmapsto} (s_0, v, \overline{u}c_{OK})$. Abhängig von der Art des letzten Schritts können zwei Fälle unterschieden werden:

1. Der letzte Schritt hat die Form $(s_0, v, A\overline{u}_1c_{OK}) \longmapsto (s_0, v, \overline{u}_0\overline{u}_1c_{OK})$ mit $\overline{u} = \overline{u}_0\overline{u}_1$ (siehe Zeile 2 der Definition von d_G). In diesem Fall ist $A ::= \overline{u}_0$ eine Regel in P, was zusammen mit der Induktionsvoraussetzung die Ableitung

$$S \overset{*}{\to} uA\overline{u}_1 \to u\overline{u}_0\overline{u}_1 = u\overline{u}$$

   ergibt.
2. Der letzte Schritt hat die Form $(s_0, xv, x\overline{u}c_{OK}) \longmapsto (s_0, v, \overline{u}c_{OK})$ (entsprechend Zeile 3 der Definition von d_G). Dann hat also u die Form $u = u_0x$ und es folgt mit der Induktionsvoraussetzung $S \overset{*}{\to} u_0x\overline{u} = u\overline{u}$. $\square$

Abschließend sei noch hinzugefügt, daß die Konstruktion auch in umgekehrter Richtung durchgeführt werden kann, d.h. es kann zu jedem Kellerautomaten eine kontextfreie Grammatik angegeben werden, die die von dem Automaten erkannte Sprache erzeugt. Zusammen mit dem im Skript Bewiesenen erhält man also, daß die Klasse der von Kellerautomaten erkannten Sprachen genau die Klasse der kontextfreien Sprachen ist:

Theorem

Eine Sprache L ist genau dann kontextfrei, wenn es einen Kellerautomaten K mit $L(K) = L$ gibt.