Grundlagen der theoretischen Informatik

Kurt Sieber

Fakultät IV, Department ETI
Universität Siegen

SS 2013

Vorlesung vom 04.06.2013

Kontextfreie Sprachen

An den Transitionen sieht man zunächst, dass *nur* die folgenden Zustandsübergänge möglich sind:

- von s mit (1) nach f oder mit (2) nach q_a,
- von q_a mit (3) nach q_a oder mit (4) nach q_b,
- von q_b mit (5) nach q_b oder mit (6) nach f.

Insbesondere ist im Zustand f *kein* Übergangsschritt mehr möglich. Also ist die gesamte Schrittfolge von der Form

- $(s, w, \varepsilon) \vdash_{(1)} (f, \varepsilon, \varepsilon)$ oder
- $$\begin{aligned}(s, w, \varepsilon) &\vdash_{(2)} (q_a, w_1, \beta_1) \vdash^m_{(3)} (q_a, w_2, \beta_2)\\ &\vdash_{(4)} (q_b, w_3, \beta_3) \vdash^n_{(5)} (q_b, w_4, \beta_4)\\ &\vdash_{(6)} (f, \varepsilon, \varepsilon)\end{aligned}$$

  mit $m, n \geq 0$

Im ersten Fall ist $w = \$ = a^0b^0\$ \in L\$$, und im zweiten Fall erhalten wir wegen der Form der Transitionen:

- $w = aw_1$ und $\beta_1 = \varepsilon$
- $w_1 = a^m w_2$ und $\beta_2 = a^m$
- $w_2 = bw_3$ und $\beta_3 = \beta_2 = a^m$
- $w_3 = b^n w_4$ mit $n \leq m$ und $\beta_4 = a^{m-n}$
- $w_4 = \$$ und $\beta_4 = \varepsilon$, also $m = n$

Insgesamt ergibt sich daraus $w = a^{n+1}b^{n+1}\$ \in L\$$. $\square$

Lemma 2.73 *Jede deterministisch kontextfreie Sprache ist kontextfrei.*

Beweis:

Wenn $L \subseteq \Sigma^*$ deterministisch kontextfrei ist, dann existiert ein Kellerautomat $M = (\Sigma \cup \{\$\}, \Gamma, Q, s, F, \Delta)$, der $L\$$ erkennt.

Kontextfreie Sprachen

Das bedeutet aber zunächst nur, dass $L\$$ kontextfrei ist. Also bleibt zu zeigen, dass auch L kontextfrei ist, d.h. dass ein Kellerautomat M' existiert, der L erkennt.

Sei $M' = (\Sigma, \Gamma, Q \times \{0,1\}, (s,0), F \times \{1\}, \Delta')$ mit

$$\begin{array}{rcll}
\Delta' & = & \{(((p,0),u,\beta),((q,0),\gamma)) \mid ((p,u,\beta),(q,\gamma)) \in \Delta\} & (1) \\
& \cup & \{(((p,0),u,\beta),((q,1),\gamma)) \mid ((p,u\$,\beta),(q,\gamma)) \in \Delta\} & (2) \\
& \cup & \{(((p,1),\varepsilon,\beta),((q,1),\gamma)) \mid ((p,\varepsilon,\beta),(q,\gamma)) \in \Delta\} & (3)
\end{array}$$

Intuition:

Immer wenn der deterministische Kellerautomat M das Ende der Eingabe mit Hilfe des Endmarkers \$ *erkennt*, dann kann der nichtdeterministische Kellerautomat M' das Ende der Eingabe *erraten*, indem er von 0 zu 1 wechselt. Nach diesem Wechsel von 0 zu 1 kann er kein Eingabezeichen mehr lesen, also kann er nur noch—mit den gleichen Übergangsschritten wie M—den Keller bearbeiten.

Es bleibt zu zeigen, dass tatsächlich $L = L(M')$ gilt.

'$\subseteq$': Sei $w \in L$, also $w\$ \in L\$ = L(M)$.

Dann existiert ein $f \in F$ mit $(s, w\$, \varepsilon) \ \vdash_M^* \ (f, \varepsilon, \varepsilon)$

Wenn wir diese Folge von Übergangsschritten dort aufteilen, wo das Ende der Eingabe (einschließlich des Endmarkers) gelesen wird, dann hat sie die Form

$$(s, w\$, \varepsilon) \ \vdash_M^* \ (p, u\$, \beta) \ \vdash_M \ (q, \varepsilon, \gamma) \ \vdash_M^* \ (f, \varepsilon, \varepsilon)$$

In der linken Teilfolge spielt der Endmarker \$ keine Rolle, weil ein Kellerautomat ein Eingabezeichen nicht "sehen" kann ohne es zu entfernen, also gilt auch

$$(s, w, \varepsilon) \ \vdash_M^* \ (p, u, \beta)$$

Aus dieser und aus den beiden rechten Teilfolgen ergibt sich dann per Definition von Δ', dass M' die folgenden Übergangsschritte machen kann:

$$((s,0),w,\varepsilon) \ \vdash^*_{(1)} \ ((p,0),u,\beta) \ \vdash_{(2)} \ ((q,1),\varepsilon,\gamma) \ \vdash^*_{(3)} \ ((f,1),\varepsilon,\varepsilon)$$

Also ist $w \in L(M')$.

'$\supseteq$': Sei $w \in L(M')$, d.h. es existiert ein $f \in F$ mit

$$((s,0),w,\varepsilon) \vdash^*_{M'} ((f,1),\varepsilon,\varepsilon)$$

Da man nur mit (2) von 0 zu 1 wechseln kann, und da man mit (3) keine Eingabe mehr lesen kann, muss diese Folge so aussehen:

$$((s,0),w,\varepsilon) \ \vdash^*_{(1)} \ ((p,0),u,\beta) \ \vdash_{(2)} \ ((q,1),\varepsilon,\gamma) \ \vdash^*_{(3)} \ ((f,1),\varepsilon,\varepsilon)$$

Daraus folgt nun umgekehrt mit der Definition von Δ'

$$(s,w\$,\varepsilon) \ \vdash^*_M \ (p,u\$,\beta) \ \vdash_M \ (q,\varepsilon,\gamma) \ \vdash^*_M \ (f,\varepsilon,\varepsilon)$$

also $w\$ \in L(M) = L\$$ und damit $w \in L$. $\square$

Ist der Endmarker wirklich notwendig in der Definition deterministisch kontextfreier Sprachen?

Im Beweis haben wir zu einem deterministischen Kellerautomaten M, der $L\$$ erkennt, einen (im allgemeinen) nichtdeterministischen Kellerautomaten M' konstruiert, der L erkennt. Das ist ein *Indiz* dafür, dass man den Endmarker benötigt.

Das folgende Beispiel zeigt, dass wir den Endmarker im allgemeinen tatsächlich nicht entbehren können.

Sei $L = \{a\}^* \cup \{a^n b^n \mid n \in \mathbb{N}\}$. Dann gilt:

- L ist deterministisch kontextfrei, d.h. es gibt einen deterministischen Kellerautomaten, der $L\$$ erkennt.
- Aber es gibt *keinen* deterministischen Kellerautomaten, der L erkennt.

Beweis:

Ein deterministischer Kellerautomat, der $L\$$ erkennt, arbeitet wie folgt:

Zunächst verschiebt er as in den Keller, so lange bis entweder \$ oder b auftaucht.

Im Falle von \$ geht er in einen Endzustand über, der ihm erlaubt, noch den Keller zu leeren.

Im Falle von b geht er in einen anderen Endzustand über, in dem er bs gegen as aufhebt und schließlich den Endmarker löscht.

Formale Definition dieses Kellerautomaten:
s. Übung

Kontextfreie Sprachen

Es bleibt zu zeigen, dass *kein* deterministischer Kellerautomat existiert, der L erkennt.

Angenommen $L = L(M)$ für einen deterministischen Kellerautomaten $M = (\Sigma, \Gamma, Q, s, F, \Delta)$.

Dann ist $a^n \in L(M)$ für jedes $n \geq 0$, also existiert für jedes $n \geq 0$ ein Zustand $f \in F$ mit $(s, a^n, \varepsilon) \vdash_M^* (f, \varepsilon, \varepsilon)$

Da es nur endlich viele Zustände $f \in F$ gibt, existieren mindestens zwei Wörter a^m und a^n mit $m \neq n$, die zur gleichen Konfiguration $(f, \varepsilon, \varepsilon)$ führen.

Nun betrachte man die Arbeitsweise von M für die beiden Eingaben $a^m b^m$ und $a^n b^m$.

Mit beiden Wörtern gelangt M zunächst in die Konfiguration $(f, \varepsilon, \varepsilon)$ und muss anschließend noch das Wort b^m lesen. Also werden beide Wörter von M gleich behandelt. Das ist ein Widerspruch zur Annahme, weil $a^m b^m \in L$ und $a^m b^n \notin L$. $\square$

Satz 2.74 *Die Klasse der deterministisch kontextfreien Sprachen ist abgeschlossen unter Komplement.*

Beweisidee:

Man vertauscht die Endzustände mit den Nicht-Endzuständen (wie beim DEA).

Das geht aber nicht gut, weil ein Kellerautomat (auch ein deterministischer) viele Möglichkeiten hat, ein Wort ‘abzulehnen’:

Er kann

- mit nichtleerem Keller halten,

- bei der Berechnung ‘steckenbleiben’, d.h. in eine Konfiguration geraten, in der das Eingabewort noch nicht abgearbeitet, aber keine Transition mehr anwendbar ist,

- in eine Endlosschleife geraten, ohne dass er das Eingabewort abgearbeitet hat.

Zum Beweis von Satz 2.74 muss man diese Möglichkeiten erst einmal ausschließen,

d.h. man muss zeigen, dass für jede deterministisch kontextfreie Sprache ein "normalisierter" Kellerautomat existiert, der seine Eingabe *immer* vollständig abarbeitet und den Keller *immer* leer macht.

Deshalb verzichten wir hier auf einen exakten Beweis. □

Wir benutzen Satz 2.74, um eine kontextfreie Sprache zu finden, die *nicht* deterministisch kontextfrei ist.

Mit anderen Worten:
Wir zeigen, dass die Klasse $\mathcal{L}_{dkf}$ der deterministisch kontextfreien Sprachen *echt* zwischen den Klassen $\mathcal{L}_{reg}$ und $\mathcal{L}_{kf}$ liegt.
(Jede reguläre Sprache ist deterministisch kontextfrei, weil man einen DEA als deterministischen Kellerautomaten auffassen kann, der den Keller nicht benutzt.)

Korollar 2.75 *Es gibt eine kontextfreie Sprache, die* nicht *deterministisch kontextfrei ist.*

Beweis:

Sei $L = \{a^l b^m c^n \mid l \neq m \text{ oder } m \neq n\}$.

L ist kontextfrei. Um dies einzusehen, kann man L in der Form

$$L = \{a^l b^m \mid l \neq m\} \circ \{c\}^* \cup \{a\}^* \circ \{b^m c^n \mid m \neq n\}$$

schreiben und für die Sprachen $\{a^l b^m \mid l \neq m\}$ und $\{b^m c^n \mid m \neq n\}$ kontextfreie Grammatiken angeben (s. Übung 6, Aufgabe 2b).

Wäre L deterministisch kontextfrei, dann wäre nach Satz 2.74 auch das Komplement $\bar{L} = \{a, b, c\}^* \setminus L$ (deterministisch) kontextfrei, und damit wäre nach Satz 2.62 auch $\bar{L} \cap \{a\}^*\{b\}^*\{c\}^*$ kontextfrei, weil $\{a\}^*\{b\}^*\{c\}^*$ regulär ist.

Aber es gilt $\bar{L} \cap \{a\}^*\{b\}^*\{c\}^* = \{a^l b^m c^n \mid l = m \text{ und } m = n\} = \{a^n b^n c^n \mid n \geq 0\}$, und diese Sprache ist *nicht* kontextfrei, wie bereits gezeigt wurde. □

Es sei noch erwähnt, dass $\mathcal{L}_{dkf}$ *nicht abgeschlossen* ist unter Vereinigung, Durchschnitt, Konkatenation oder Kleene-Abschluss.

Das Gegenbeispiel für den Durchschnitt ist das gleiche wie für $\mathcal{L}_{kf}$ (Korollar 2.69). Man wähle $L_1 = \{a^n b^n \mid n \geq 0\} \circ \{c\}^*$ und $L_2 = \{a\}^* \circ \{b^n c^n \mid n \geq 0\}$. Beide Sprachen sind deterministisch kontextfrei, denn man kann für sie ähnliche deterministische Kellerautomaten konstruieren wie für die Sprache $\{a^n b^n \mid n \geq 0\}$.

Aber $L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 0\}$ ist—wie wir bereits wissen—noch nicht einmal kontextfrei.

Es folgt sofort, dass $\mathcal{L}_{dkf}$ auch unter Vereinigung nicht abgeschlossen sein kann, weil man ja den Durchschnitt zweier Mengen nach den De Morganschen Regeln durch eine Kombination von Komplement- und Vereinigungsbildung erhalten kann.