Grundlagen der theoretischen Informatik

Kurt Sieber

Fachbereich Mathematik/Theoretische Informatik
Universität Siegen

Vorlesung vom 23.11.2004 (Stand: 24.11.2004)

Wie beweist man, dass eine KFG die gewünschte Sprache erzeugt?

Beispiel:

Sei $L = \{a^n b^n \mid n \geq 0\}$
und $G = (\{a, b\}, \{S\}, S, P)$ mit $P = \{S \to aSb, S \to \varepsilon\}$

Behauptung: $L(G) = L$.

Beweis:

'$\subseteq$': Sei $w \in \Sigma^*$ mit $S \overset{*}{\Rightarrow} w$.

Es ist zu zeigen, dass $w \in L$.

Dazu beweist man eine 'passende Behauptung' durch *Induktion über die Länge der Ableitung*.

Für den Induktionsschritt hat man zwei Möglichkeiten: Entweder man spaltet den *ersten* Ableitungsschritt ab oder den *letzten*.

Will man den *letzten* Ableitungsschritt abspalten, so muss man eine Behauptung für alle $u \in (N \cup \Sigma)^*$ mit $S \overset{*}{\Rightarrow} u$ aufstellen, z.B.:

$$\text{Wenn } u \notin \Sigma^*, \text{ dann ist } u = a^n S b^n \text{ für ein } n \geq 0. \qquad (*)$$

$S \overset{0}{\Rightarrow} u$:

Dann ist $u = S = a^0 S b^0$.

$S \overset{+}{\Rightarrow} u$, d.h. $S \overset{*}{\Rightarrow} v \Rightarrow u$ für ein $v \notin \Sigma^*$:

Dann ist $v = a^n S b^n$ nach Induktionsannahme. Wegen $u \notin \Sigma^*$ kann im Ableitungsschritt $v \Rightarrow u$ nur die Produktion $S \to aSb$ angewandt worden sein, also ist $u = a^{n+1} S b^{n+1}$.

Man kann die Argumentation noch verkürzen: Es genügt offensichtlich zu zeigen, dass (*) für S gilt und bei jedem Ableitungsschritt erhalten bleibt. (*) ist also eine *Invariante* für die Ableitungen aus S.

Aus (*) folgt schließlich $L(G) \subseteq L$:

Wenn $w \in L(G)$, dann gilt $S \stackrel{*}{\Rightarrow} u \Rightarrow w$ für ein $u \notin \Sigma^*$, also $u = a^n S b^n$ für ein $n \geq 0$. Wegen $w \in \Sigma^*$ kann $u \Rightarrow w$ nur mit der Produktion $S \to \varepsilon$ erfolgen, also ist $w = a^n b^n \in L$.

Die Abspaltung des *ersten* Ableitungsschrittes ist in unserem Falle wesentlich einfacher.

Man stellt eine Behauptung für alle $w \in \Sigma^*$ mit $S \stackrel{*}{\Rightarrow} w$ auf, nämlich:

$$w \in L \qquad (**)$$

$S \stackrel{1}{\Rightarrow} w$:

Dann ist $w = \varepsilon \in L$.

$S \Rightarrow aSb \stackrel{*}{\Rightarrow} w$:

Dann ist $w = avb$ mit $S \stackrel{*}{\Rightarrow} v$, also gilt $v \in L$, d.h. $v = a^n b^n$ für ein $n \geq 0$ und damit $w = a^{n+1} b^{n+1}$.

Geht es immer so einfach?

Im allgemeinen muss man für *jedes* $A \in N$ eine Behauptung über alle $w \in \Sigma^*$ mit $A \stackrel{*}{\Rightarrow} w$ aufstellen, und diese Behauptungen durch *simultane Induktion* beweisen.

Mit anderen Worten: Wenn $|N| = k$, so muss man eine Behauptung über die k Sprachen $L_A(G) = \{w \in \Sigma^* \mid A \stackrel{*}{\Rightarrow} w\}$ aufstellen und durch simultane Induktion beweisen.

'$\supseteq$': Sei $w \in L = \{a^n b^n \mid n \geq 0\}$.

Durch Induktion über $|w|$ beweist man $w \in L(G)$.

$|w| = 0$, d.h. $w = \varepsilon$:

Dann gilt $S \Rightarrow w$ mit Produktion $S \rightarrow \varepsilon$.

$|w| > 0$, d.h. $w = a^n b^n$ mit $n > 0$:

Dann ist $w = aa^{n-1}b^{n-1}b$, also $w = avb$ mit $v \in L$. Nach Induktionsannahme gilt $S \stackrel{*}{\Rightarrow} v$, also $S \Rightarrow aSb \stackrel{*}{\Rightarrow} avb = w$.

Auch diese Richtung wird schwieriger, wenn man mehrere Nichtterminalzeichen hat.

Dann kann man wieder eine geeignete Behauptung über die Sprachen $L_A(G)$ aufstellen und durch Induktion beweisen.

Wir haben im Beispiel die folgenden Eigenschaften von Ableitungen benutzt:

Lemma 2.10

1. *Für alle* $A \in N$ *und* $u, v, w \in (N \cup \Sigma)^*$ *gilt:*

   Wenn $A \overset{*}{\Rightarrow} v$, *dann* $uAw \overset{*}{\Rightarrow} uvw$.

2. *Für alle* $A \in N$ *und* $u, w, x \in \Sigma^*$ *gilt:*

   Wenn $uAw \overset{*}{\Rightarrow} x$, *dann existiert ein* $v \in \Sigma^*$ *mit* $A \overset{*}{\Rightarrow} v$ *und* $x = uvw$.

Weiteres Beispiel:

Sei $L = \{w \in \{a,b\}^* \mid \#_a(w) = \#_b(w)\}$.

Es gibt eine einfache Grammatik, die L erzeugt, nämlich:

$G = (\{a,b\},\{S\},S,P)$ mit $P = \{S \to SS, S \to aSb, S \to bSa, S \to \varepsilon\}$

Behauptung: $L(G) = L$.

Beweis:

'$\subseteq$': Für jedes $u \in (N \cup \Sigma)^*$ mit $S \overset{*}{\Rightarrow} u$ gilt:

$$\#_a(u) = \#_b(u) \qquad (*)$$

denn (*) gilt für S und bleibt bei jedem Ableitungsschritt erhalten (weil in jeder Produktion gleich viele as und bs hinzukommen).

Also gilt (*) insbesondere für jedes $w \in L(G)$, d.h. $L(G) \subseteq L$.

'$\supseteq$': Sei $w \in L$, $w = a_1 \dots a_n$.

Wenn $w = \varepsilon$, so gilt $S \Rightarrow w$ mit Produktion $S \rightarrow \varepsilon$.

Es bleibt der Fall $|w| > 0$ zu betrachten.

Sei $f : \{a, b\}^* \rightarrow \mathbb{Z}$ mit $f(v) = \#_a(v) - \#_b(v)$.

Dann gilt $L = \{v \in \{a, b\}^* \mid f(v) = 0\}$.

Wir betrachten den 'Verlauf' der Funktion f für die Präfixe von w, d.h. für die Wörter $w_i = a_1 \dots a_i$ mit $0 \leq i \leq n$.

Es gilt $f(w_0) = f(\varepsilon) = 0$ und $f(w_n) = f(w) = 0$, weil $w \in L$.

1. Fall: $f(w_i) \geq 0$ für $i = 0, \dots, n$

Dann ist $f(w_1) = 1$ und $f(w_{n-1}) = 1$, also $w = avb$ für ein $v \in L$.

Nach Induktionsannahme gilt $S \overset{*}{\Rightarrow} v$, also $S \Rightarrow aSb \overset{*}{\Rightarrow} avb = w$.

2. Fall: $f(w_i) \leq 0$ für $i = 0, \ldots, n$

Analog zum 1. Fall folgt $w = bva$ für ein $v \in L$,

also gilt $S \Rightarrow bSa \overset{*}{\Rightarrow} bva = w$.

3. Fall: Es existieren $i, j \in \{0, \ldots, n\}$ mit $f(w_i) > 0$ und $f(w_j) < 0$.

Dann existiert eine Zahl k zwischen i und j mit $f(w_k) = 0$, also $w_k = a_1 \ldots a_k \in L$.

Dann ist auch $w'_k = a_{k+1} \ldots a_n \in L$ und es gilt $w = w_k w'_k$.

Nach Induktionsannahme gilt $S \overset{*}{\Rightarrow} w_k$ und $S \overset{*}{\Rightarrow} w'_k$, also $S \Rightarrow SS \overset{*}{\Rightarrow} w_k S \overset{*}{\Rightarrow} w_k w'_k = w$. $\square$

Die Beispiele zeigen, dass es kontextfreie Sprachen gibt, die *nicht* regulär sind.

Wir beweisen jetzt: Jede reguläre Sprache ist kontextfrei.

Um reguläre Sprachen zu erzeugen, reichen sehr spezielle Grammatiken aus.

Definition 2.11 *Eine Grammatik heißt* rechtslinear, *wenn jede Produktion von der Form $A \to aB$ oder $A \to \varepsilon$ (mit $A, B \in N$ und $a \in \Sigma$) ist. Sie heißt* linkslinear, *wenn jede Produktion von der Form $A \to Ba$ oder $A \to \varepsilon$ ist.*

In der Literatur sind oft noch andere Produktionen zugelassen, etwa $A \to a$, $A \to B$ oder $A \to wB$ mit $w \in \Sigma^*$. Dadurch werden die Grammatiken nicht mächtiger.

Beispiel:

Sei $L = \{a^{2n} \mid n \geq 0\}$
und $G = (\{a\}, \{S_0, S_1\}, S_0, P)$ mit $P = \{S_0 \to aS_1, S_1 \to aS_0, S_0 \to \varepsilon\}$

Behauptung: $L(G) = L$.

Beweis:

'$\subseteq$': Jedes aus S_0 ableitbare Wort, das nicht in $\{a\}^*$ liegt, hat die Form $a^{2n}S_0$ oder $a^{2n+1}S_1$. (Induktion über die Länge der Ableitung).

Da jedes $w \in L(G)$ nur mit Produktion $S_0 \to \varepsilon$ aus einem solchen Wort entstehen kann, muss $w = a^{2n}$ sein für ein $n \geq 0$.

'$\supseteq$': Für jedes $n \geq 0$ gilt

$$S_0 \Rightarrow aS_1 \Rightarrow aaS_0 \Rightarrow \ldots \Rightarrow a^{2n}S_0 \Rightarrow a^{2n} \qquad \square$$

Am Beispiel sieht man, dass die Nichtterminalzeichen einer rechtslinearen Grammatik eine ähnliche Rolle spielen wie die Zustände eines endlichen Automaten.

In jedem Ableitungsschritt (außer dem letzten) wird ein Terminalzeichen erzeugt und das Nichtterminalzeichen am Ende des Wortes wird eventuell verändert.

So kann man sich im Nichtterminalzeichen eine Information über die bereits erzeugten Terminalzeichen merken.

Im Beispiel: S_0 bedeutet, dass bisher eine gerade Anzahl von as erzeugt wurde, S_1 bedeutet, dass eine ungerade Anzahl von as erzeugt wurde.

Satz 2.12

1. *Zu jedem NDEA A kann man eine rechtslineare Grammatik G mit $L(G) = L(A)$ konstruieren.*
2. *Zu jeder rechtslinearen Grammatik G kann man einen NDEA A mit $L(A) = L(G)$ konstruieren.*

Beweis:

1. Sei $A = (\Sigma, Q, s, F, \Delta)$ ein NDEA.

   Wir dürfen annehmen, dass $\Sigma \cap Q = \emptyset$.

   Sei $G = (\Sigma, Q, s, P)$ mit

   $P = \{p \to aq \mid (p, a, q) \in \Delta\} \cup \{p \to \varepsilon \mid p \in F\}$

   Dann folgt durch eine einfache Induktion über $|w|$:

$$p \overset{*}{\Rightarrow}_G wq \;\Leftrightarrow\; (p, w) \vdash (q, \varepsilon) \qquad (*)$$

Also gilt

$$
\begin{array}{rcl}
w \in L(G) & \Leftrightarrow & s \stackrel{*}{\Rightarrow}_G w \\
& \Leftrightarrow & \text{es existiert ein } p \in Q \text{ mit } s \stackrel{*}{\Rightarrow}_G pw \Rightarrow w \\
& & \text{und } (p \to \varepsilon) \in P \\
& & \text{(weil der letzte Ableitungsschritt nur mit einer} \\
& & \;\varepsilon\text{-Produktion erfolgen kann)} \\
& \Leftrightarrow & \text{es existiert ein } p \in Q \text{ mit } (s, w) \vdash_A^* (p, \varepsilon) \\
& & \text{und } p \in F \\
& & \text{(wegen (*) und der Definition von } P\text{)} \\
& \Leftrightarrow & w \in L(A)
\end{array}
$$

und damit ist $L(G) = L(A)$ bewiesen.

2. Sei $G = (\Sigma, N, S, P)$ eine rechtslineare Grammatik.

   Wir definieren $A = (\Sigma, N, S, F, \Delta)$ mit

   $F = \{B \in N \mid (B \to \varepsilon) \in P\}$ und

   $\Delta = \{(B, a, C) \in N \times \Sigma \times N \mid (B \to aC) \in P\}$

   Dann folgt durch eine einfache Induktion über $|w|$:

$$B \overset{*}{\Rightarrow}_G wC \iff (B, w) \vdash_A^* (C, \varepsilon)$$

   und ähnlich wie oben ergibt sich $L(A) = L(G)$ $\square$

Damit haben wir eine weitere Charakterisierung regulärer Sprachen.

Korollar 2.13 *Eine Sprache ist genau dann regulär, wenn sie sich von einer rechtslinearen Grammatik erzeugen lässt.*

Außerdem ist damit natürlich beweisen, dass jede reguläre Sprache kontextfrei ist, also gilt

Satz 2.14 $\mathcal{L}_{reg} \subseteq \mathcal{L}_{kf}$

wobei $\mathcal{L}_{kf}$ die Klasse der kontextfreien Sprachen bezeichnet.

Unsere Beispiele zeigen, dass diese Inklusion echt ist, wenn das Alphabet Σ mindestens zwei Elemente enthält. (Bei einelementigem Alphabet stimmen die beiden Sprachklassen überein.)