Vorlesungsmitschrift zur Vorlesung Theoretische Informatik

Lutz Dittrich

Stefan Liske

15. Juni 2000

Zwischenbetrachtung / Resümee:
Wir kennen bereits den Mechanismus des Erkennens (durch Automaten) einer Sprache (erzeugt über einem endlichen Alphabet). Diese Sprachen, losgelöst von den Akzeptoren, genauer zu betrachten und so den Mechanismus des Erzeugen (durch Grammatiken oder reguläre Ausdrücke) kennenzulernen ist Aufgabe der folgenden Absätze.

Sprachen wurden von Noam Chomsky (*1928) genauer untersucht. Von ihm stammt die Unterteilung der Klasse aller Sprachen in vier Kategorien:

Sprachen der Stufe:	sind:	und werden akzeptiert von:
3	aufzählbare Sprachen	Turingmaschinen
2	kontextsensitive Sprachen	linear beschränkten Automaten
1	kontextfreie Sprachen	Push-Down-Automaten / Kellerautomaten
0	reguläre Sprachen	deterministischen Automaten

Unser erstes Ziel ist es, die Sprachklasse zu klassifizieren, die wir durch endliche Akzeptoren erkennen. Es wird gezeigt werden, daß dies genau der Klasse der regulären Sprachen entspricht:

4.3. Reguläre Sprachen

Definition N: (Syntaxdefinition)
Sei X ein Alphabet und V={ (,) , * , ∪ , • , ∅ } ein Vorrat an Zeichen.
Dann ist die Mengen der ***regulären Ausdrücke*** Reg(X) folgendermaßen definiert:

i) $\forall x\in X(\, x\in Reg(X)\,) \Leftrightarrow X\subseteq Reg(X)$
ii) $\emptyset\in Reg(X)$
iii) $\forall R,Q\in Reg(X)(\,(R\cup Q)\in Reg(X) \wedge (R\bullet Q)\in Reg(X)\,)$
iv) $\forall R\in Reg(X)(\, R^*\in Reg(X)\,)$
v) Reg(X) ist kleinste Teilmenge von $(X\cup V)^*$, die i) bis iv) erfüllt

Beispiele:
Sei $X=\{x_1,x_2,x_3\}$ Dann sind u.a. folgende Ausdrücke reguläre Ausdrücke:
- x_1
- $(((x_2\bullet x_1) \bullet x_3)\cup x_3)$
- $((x_1\bullet x_2)^*\cup x_1)^*$

Diese Folgen, bestehend aus Zeichen und Elementen aus $X\cup V$ *sind* bisher *jedoch sinnleer*! Sie entstehen durch bloßes Anwenden der Regeln i) bis iv). Die nun notwendige Interpretation folgt jetzt:

Definition O: (Semantikdefinition)
Die Bedeutung eines regulären Ausdruckes über X wird durch folgende Abbildung $\delta:Reg(X)\rightarrow 2^{X^*}$ festgelegt, die wie folgt induktiv definiert ist:

i) $x\in X \Rightarrow \delta(x)=\{x\}$
ii) $\delta(\emptyset)=\emptyset$
iii) $\forall R,Q\in Reg(X)(\, \delta(R\cup Q)=\delta(R)\cup\delta(Q) \wedge \delta(R\bullet Q)=\delta(R)\bullet\delta(Q)\,)$
iv) $\forall R\in Reg(X)(\, \delta(R^*)=\delta(R)^*\,)$

Beispiele:
- $\delta(\,\emptyset\,) = \emptyset$
  dem sinnleeren Zeichen „∅“ wird die leere Menge zugeordnet;
- $\delta(\,\emptyset^*\,) = \delta(\emptyset)^*=\emptyset^*=\{\varepsilon\}$
  der sinnleeren Zeichenfolge „∅*“ wird nach Regel iv) der Semantikdefinition die mengentheoretische *-Bildung von „δ(∅)“ zugewiesen, was seinerseits wiederum nach Regel ii) die leere Menge ist. Nach Definition der *-Bildung folgt: $\emptyset^*=\{\varepsilon\}$;

Bemerkung:
ergänzend zu den Semantikregeln gibt es Prioritätsregeln, die eine Änderung der Syntax mit sich bringen, jedoch die Bedeutung des Ausdruckes unverändert lassen:

i) „•“ wird vor „∪“ interpretiert
ii) Weglassen äußerer Klammern
iii) Weglassen des Zeichens „•“

Beispiel:

〈 „$(x_1 \cup (x_2 \bullet x_3))$" entspricht „$x_1 \cup (x_2 \bullet x_3)$" nach Prioritätsregel ii)
entspricht „$x_1 \cup x_2 \bullet x_3$" nach Prioritätsregel i) und ii)
entspricht „$x_1 \cup x_2 x_3$" nach Prioritätsregel iii)

Definition P:

Sei X ein Alphabet.
Eine Sprache $L \subseteq X^*$ heißt reguläre Sprache, wenn es einen regulären Ausdruck $R \in Reg(X)$ mit $\delta(R)=L$ gibt. ($L \subseteq X^*$ heißt reguläre Sprache $:\Leftrightarrow \exists R \in Reg(X)(\ \delta(R)=L\ )$.
Die Menge aller regulären Sprachen über X bezeichnen wir mit R(X).
Die Klasse aller regulären Sprachen bezeichnen wir mit R.

Äquivalenzproblem für reguläre Ausdrücke:

Sei R ein regulärer Ausdruck mit: $R = a(a^* \cup b^*)^* \cup (b \cup c)(a \cup b \cup c)^* \cup (ca)^*$,
dann folgt für $\delta(R)$: $\delta(R) = \{\ w \in X^* \mid$ $w \in \{a,b\}^*$ und w beginnt mit einem a
oder $w \in \{a,b,c\}^*$ und w beginnt nicht mit a
oder $w=\varepsilon$
$\}$

Der dritte Fall $(ca)^*$ kann auf $w=\varepsilon$ „runtergewertet" werden; Dies ergibt sich aus: $(ca)^* \setminus \{\varepsilon\} \subseteq (b \cup c)(a \cup b \cup c)^*$, somit muß nur noch der Fall $\{\varepsilon\}$ hinzugefügt werden.

Sei R' ein regulärer Ausdruck mit: $R' = a(a \cup b)^* \cup (b \cup c)(a \cup b \cup c)^* \cup \emptyset^*$
Ersichtlich ist: $\delta(R) = \delta(R')$

Aus dieser Betrachtung ergibt sich folgende
<u>Frage:</u> Gibt es einen Algorithmus, der für jedes beliebige Alphabet X und für je zwei reguläre Ausdrücke $R,R' \in Reg(X)$ entscheidet, ob $\delta(R) = \delta(R')$?
<u>Antwort:</u> Ja, Dieses Problem ist algorithmisch lösbar, aber sehr aufwendig (NP-hart [nichtdeterministisch in polynomiel viel Zeit lösbar]).

Satz Q: (Gleichheit von R und DA)

Die Klasse aller regulären Sprachen und die Klasse aller, durch endliche Automaten akzeptierbarer Sprachen ist gleich. (R=DA)

Beweis von Satz Q:

Zeige zuerst die mengentheoretische Inklusion von R in DA („⊆"):

Definition N $\Rightarrow$ R ist kleinste Klasse von Sprachen, die
die leere Menge enthält,
die alle Mengen $\{x\}$ (mit $x \in X$) enthält,
die gegen Vereinigung, Konkatenation und -Bildung abgeschlossen ist;
Lemma E $\Rightarrow$ DA enthält die leere Menge, DA enthält alle endlichen Mengen;
Satz D, Satz K $\Rightarrow$ DA ist abgeschlossen gegen -Bildung, Konkatenation und Vereinigung;
$\Rightarrow R \subseteq DA$

Zeige nun die mengentheoretische Inklusion von DA in R („⊇"):

Beweisidee: Wähle einen beliebigen endlichen deterministischen Akzeptor (der L(A) erkennt) und konstruiere dazu einen regulären Ausdruck R (mit der Bedeutung $\delta(R)$), so daß $L(A)=\delta(R)$.
Sei $A=(X,S,\delta,1,F)$ ein endlicher deterministischer Akzeptor mit n Zuständen (o.B.d.A.: $S=\{1,...,n\}$).
Beweismethode: Vollständige Induktion über die (Nummern der) Zwischenzustände, die man benötigt, um von einem Zustand i in einen Zustand j zu gelangen.
Definiere für alle $i,j \in S$ und $k \in \{0,1,...,n\}$ (entspricht: $k \in (\{0\} \cup S)$) :
$W_{ij}^k := \{\ w \in X \mid \delta(i,w)=j \wedge \forall u,v \in X^* \setminus \{\varepsilon\}(\ w=uv \Rightarrow \delta(i,u) \leq k\ )\ \}$
(W_{ij}^k sei die Menge aller Wörter mit denen man vom Zustand i in den Zustand j kommt und unterwegs nur über die Zwischenzustände von 1 bis k kommt)

<u>Induktionsanfang:</u> (k=0)
$W_{ij}^0 \subseteq X \cup \{\varepsilon\}$ (von i nach j ohne Zwischenzustände $\Rightarrow |w|=1$ oder $w=\varepsilon$) $\left\{ \begin{array}{l} X \\ \emptyset^* \end{array} \right.$
W_{ij}^0 kann also durch einen regulären Ausdruck dargestellt werden, nämlich:

<u>Induktionsannahme:</u> (k ist eine beliebiges natürliche Zahl)
Wir nehmen an, für ein beliebiges $k \geq 0$ und für alle $i,j \in S$ seien die Mengen W_{ij}^k bereits konstruiert und durch einen regulären Ausdruck darstellbar. Dann gilt für alle $i,j \in S$:

<u>Induktionsschritt:</u> (zeige $\forall i,j \in \{1,...,n\} \forall k \in \{0,1,...,n\} (W_{ij}^{k} \in R \Rightarrow W_{ij}^{k+1} \in R)$)
Es zeigt sich: $W_{ij}^{k+1} = W_{ij}^{k} \cup W_{i(k+1)}^{k} \bullet (W_{(k+1)(k+1)}^{k})^* \bullet W_{(k+1)j}^{k}$
Die Menge der Wege von i nach j über die Zwischenzustände 1 bis (k+1) (W_{ij}^{k+1}) setzt sich zusammen aus (trivialerweise) allen Wegen von i nach j über die ersten (k) Zustände (W_{ij}^{k}) UND den Wegen, die tatsächlich den Zustand (k+1) passieren.
Diese Wege kann man als Wegfolgen (Konkatenation einzelner Wegstücke) betrachten, wobei man für den Anfang jeder Wegfolge der Zustand i, für das Ende der Zustand j angenommen wird (es geht ja von i nach j). Zwischendurch jedoch wähle man die Wegstücke so, so daß sie mit dem Zustand (k+1) enden und beginnen. Also vom Zustand i in den Zustand (k+1) gehen (erstes Wegstück), diesen Zustand (k+1) eventuell immer wieder durchlaufen (weitere Wegstücke) und anschließend in den Zustand j gehen (letztes Wegstück) ($W_{i(k+1)}^{k} \bullet (W_{(k+1)(k+1)}^{k})^* \bullet W_{(k+1)j}^{k}$).
Die einzelnen Wegstücke haben nur Zwischenstops in Zuständen von 1 bis k.
Und diese Teilstücke (W_{ij}^{k} , $W_{i(k+1)}^{k}$, $(W_{(k+1)(k+1)}^{k})^*$ und $W_{(k+1)j}^{k}$) sind bereits (nach Induktionsannahme) als regulärer Ausdruck dargestellt. Da auch die Vereinigung, die *-Bildung und die Konkatenation wiederum reguläre Ausdrücke ergeben, ist auch die Folge ($W_{ij}^{k} \cup W_{i(k+1)}^{k} \bullet (W_{(k+1)(k+1)}^{k})^* \bullet W_{(k+1)j}^{k}$) wiederum ein regulärer Ausdruck.

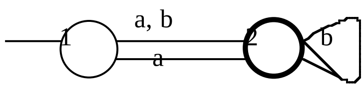
<u>Fazit:</u> Ist W_{ij}^{k} als regulärer Ausdruck darstellbar, dann ist auch W_{ij}^{k+1}!
Da wir den Anfang mit k=0 gezeigt haben gilt dies folglich für alle beliebigen natürlichen Zahlen.

Die akzeptierte Sprache des Akzeptors L(A) ist folglich die Vereinigung aller Wege vom Startzustand in alle möglichen Endzustände (wobei als Zwischenzustände alle Zustände des Akzeptors möglich sind):
$L(A) = \bigcup_{j \in F} W_{1j}^{n}$; dies ist ebenfalls ein regulärer Ausdruck

$\Rightarrow L(A) \in R$, für einen beliebigen endlichen deterministischen Akzeptor A
$\Rightarrow DA \subseteq R$

Dieser Beweis hat einen schönen Nebeneffekt. Er gibt eine Anleitung, um für jedem Akzeptor die erkannte Sprache zu ermitteln, indem wir den zugehörigen regulären Ausdruck bestimmen.

<u>Beispiel:</u>
Gegeben sei der Akzeptor $A = (\{a,b\},\{1,2\},\delta,1,\{2\})$ mit folgendem (graphisch dargestellten) δ:



Dann ist ersichtlich (Betrachtung 1):

$W_{11}^{0} = \{\varepsilon\}$ $\quad W_{12}^{0} = \{a,b\}$
$W_{21}^{0} = \{b\}$ $\quad W_{22}^{0} = \{\varepsilon,b\}$

daraus folgt (Betrachtung 2):

$W_{11}^{1} = W_{11}^{0} \cup W_{11}^{0} \bullet (W_{11}^{0})^* \bullet W_{11}^{0} = \{\varepsilon\}$
$W_{12}^{1} = W_{12}^{0} \cup W_{11}^{0} \bullet (W_{11}^{0})^* \bullet W_{12}^{0} = \{a,b\}$
$W_{21}^{1} = W_{21}^{0} \cup W_{21}^{0} \bullet (W_{11}^{0})^* \bullet W_{12}^{0} = \{a\}$
$W_{22}^{1} = W_{22}^{0} \cup W_{21}^{0} \bullet (W_{11}^{0})^* \bullet W_{12}^{0} = \{\varepsilon,b\} \cup \{aa,ab\}$

daraus folgt (Betrachtung 3):

$$\begin{aligned}
W_{11}^{2} = W_{11}^{1} \cup W_{12}^{1} \bullet (W_{22}^{1})^* \bullet W_{21}^{1} &= \{\varepsilon\} \cup \{a,b\}\{\varepsilon,b,aa,ab\}^*\{a\} \\
&= \{\varepsilon\} \cup \{a,b\}\{b,aa,ab\}^*\{a\} \\
W_{12}^{2} = W_{12}^{1} \cup W_{12}^{1} \bullet (W_{22}^{1})^* \bullet W_{22}^{1} &= \{a,b\} \cup \{a,b\}\{\varepsilon,b,aa,ab\}^*\{\varepsilon,b,aa,ab\} \\
&= \{a,b\} \cup \{a,b\}\{\varepsilon,b,aa,ab\}^* \\
&= \{a,b\} \cup \{b,aa,ab\}^* \\
W_{21}^{2} = W_{21}^{1} \cup W_{22}^{1} \bullet (W_{22}^{1})^* \bullet W_{21}^{1} &= \{a\} \cup \{\varepsilon,b,aa,ab\}\{\varepsilon,b,aa,ab\}^*\{a\} \\
&= \{a\} \cup \{b,aa,ab\}^* \\
W_{22}^{2} = W_{22}^{1} \cup W_{22}^{1} \bullet (W_{22}^{1})^* \bullet W_{22}^{1} &= \{\varepsilon,b,aa,ab\} \cup \{\varepsilon,b,aa,ab\}\{\varepsilon,b,aa,ab\}^*\{\varepsilon,b,aa,ab\} \\
&= \{b,aa,ab\}^*
\end{aligned}$$

nun ergibt sich für die Betrachtung von L(A):

$$L(A) = \bigcup_{j \in F} W_{1j}^{2} = W_{12}^{2} = \{a,b\} \cup \{b,aa,ab\}^*$$

ein entsprechender regulärer Ausdruck R ist demnach:

$$\begin{aligned}
R &= (a \cup b) \bullet ((b \cup (a \bullet a)) \cup (a \bullet b))^* \\
&= (a \cup b)((b \cup aa) \cup ab)^*
\end{aligned}$$