Script zur Lehrveranstaltung

Automatentheorie und formale Sprachen

Prof. Annika Wagner

Florian Schleich

April 12, 2013

Hochschule Fulda
University of Applied Sciences

1 Grammatiken

In diesem Kapitel lernen wir, wie man Grammatiken definiert und Wörter aus diesen Grammatiken erzeugt.

1.1 Symbole

Symbole sind alle verwendbaren Zeichen um durch Kombination Wörter zu erschaffen.

1.2 Alphabet

Ein Alphabet, abgekürzt Σ, ist eine endliche Menge von einzelnen Symbolen. Ein Alphabet wird als endliche Menge von Symbolen definiert.

$\Sigma = \{☺, a, Anton\}$

Neben einzelnen Buchstaben können auch Wörter im Sinne der deutschen Sprache verwendet werden. Es können aber auch andere Symbole (☺) verwendet werden. Bei Verwendung von Wörtern als Symbole eines Alphabets ist aber zu beachten, dass bei der späteren Verwendung die einzelnen Buchstaben oder Teile des Wortes nicht als Symbole genommen werden dürfen, sondern das gesamte Wort als ein Symbol bezeichnet wird. Im hier verwendeten Fall darf also das große A nicht als Symbol verwendet werden darf. ebenso dürfen auch keine Wortteile wie *ton* nicht verwendet werden. Ein zulässiges Symbol ist allerdings das gesamte Wort *Anton*.

1.3 Wörter

Wörter werden aus Kombination von einzelnen Symbolen gebildet. Die bildbaren Wörter werden in einer Menge $\Sigma*$ gesammelt. Wird die Menge aller Wörter gebildet muss jetzt explizit das leere Wort ϵ angegeben werden. Auf Grund der in der Grammatik verwendeten Regeln wird die Menge der möglichen Wörter eingeschränkt. In unserem Beispiel kann die Menge aller Wörter so aussehen:

$\Sigma* = \{\epsilon, \smiley, a, Anton, \smiley a, aAnton, \smiley Anton, a\smiley, Antona, Anton\smiley, \smiley\smiley, ...\}$

Die Menge wird gebildet, in dem erst das leere Wort, dann alle einstelligen Wörter, alle zweistelligen Wörter, alle dreistelligen Wörter usw. in die Menge integriert werden. Die Menge der Wörter ist unendlich, da die Länge der Wörter nicht bestimmt ist.

Die Länge eines Wortes $|w|$ bezeichnet die Anzahl der für das Wort verwendeten Symbole.

$|\epsilon| = 0$
$|a| = 1$
$|Anton| = 1$
$|aa| = 2$

Figure 1.1: Beispiele für Wortlängen

1.4 Grammatiken

Eine Grammatik G wird definiert durch:

- eine Menge von Variablen $V = \{S, U\}$, die bei der Ableitung verschwinden,
- eine Menge von Terminalsymbolen Σ, gelichbedeutend mit dem definierten Alphabet
- ein Startsymbol $S \in V$
- und eine Menge von Regeln P

Formal wird eine Grammatik also beschrieben als:

$G = (V, \Sigma, S, P)$

1.4.1 Regeln

Eine Regel besteht aus einer linken und einer rechten Regelseite. Bei der Ableitung von Regeln wird der Ausdruck auf der linken Seite durch den Ausdruck auf der rechten Seite ersetzt.

Bei der Regel $S \rightarrow Anton$ wird die Variable S durch das Symbol *Anton* ersetzt.

1.4.2 Ableitung

Bei der Ableitung von Regeln beginnt man beim Startsymbol und ersetzt das Startsymbol durch den entsprechenden Ausdruck einer Regel, die auf das Startsymbol passt. Zwischen zwei Ableitungsschritten kommt dann ein Doppelpfeil ($\Rightarrow$).

$S \rightarrow Anton$ **(Regel)**
$S \Rightarrow Anton$ **(Ableitung)**

Figure 1.2: Einfaches Beispiel für eine Ableitung

1.4.3 Regelalternativen

Regeln können auch Alternativen enthalten, um mehr Möglichkeiten zu bieten Wörter zu erzeugen.

Alternativen werden auf der rechten Regelseite durch ein Pipe-Symbol (|) eingeleitet. Eine Regel mit Alternativen könnte folgendermaßen aussehen:

$S \rightarrow Anton|Anton☺|Antona$

Dadurch wären die Wörter *Anton*, *Anton*☺ und *Antona* für die Variable S möglich.

1.4.4 Rekursion in Regeln

$S \rightarrow AntonG$
$G \rightarrow$ ☺ $|$ ☺ G

Figure 1.3: Grammatik mit Rekursion

Rekursionen in Grammatiken können Wörter erzeugen, deren Länge (Betrag) unendlich lang sind. Eine mögliche Ableitung für die obige Grammatik könnte folgendermaßen lauten.

$S \Rightarrow AntonG$
$\Rightarrow Anton$☺G
$\Rightarrow Anton$☺☺

Figure 1.4: Ableitung für eine rekursive Grammatik

Wir sehen also, dass durch die Rekursion auf G die Wörter unendlich lang werden können und durch das Symbol ☺ aufgefüllt werden. Des weitern stellen wir fest, dass Ableitungen erst enden, wenn alle Variablen entfernt, bzw. ersetzt wurden.

1.4.5 Mengen von Regeln

Als weitere Komponente für die Definition von Grammatiken wird die Menge aller Regeln der Grammatik benötigt. Im vorherigen Abschnitt haben wir gelernt, wie Regeln definiert werden. Die Menge der Regeln einer Grammatik ist eine Teilmenge des kartesische Produktes der Elemente aus V und Σ. Das kartesische Produkt wird durch die Kombination aller Elemente einer Menge mit sämtlichen Elementen einer anderen Menge gebildet.

$A = \{0, 1\}$
$B = \{x\}$
$A \times B = \{(0, x), (1, x)\}$

Figure 1.5: Kartesisches Produkt der Mengen A und B

Die Menge aller Regeln einer Grammatik wird nun durch $(V \cup \Sigma)^* \times (V \cup \Sigma)^*$ gebildet. Durch diese Definition lassen sich sämtliche Kombinationen aus Terminalsymbolen und Variablen auf beden Seiten der Regeln ausfstellen. Es ist aber meistens gewünscht nur ein paar bestimmte Regeln zuzulassen. Man bildet also eine bestimmte Teilmenge der eben gebildeten Gesamtmenge.

1.5 Kontextfrei oder regulär

Kontextfreie Grammatiken haben auf der linken Regelseite genau **eine** Variable und eine Beliebige Kombination aus Terminalsymbolen und Variablen auf der rechten Regelseite. Grammatiken sind regulär, wenn die verwendeten Regeln auf der linken Regelseite **eine** Variable und auf der rechten Seite entweder das leere Wort (ϵ), genau ein Terminalsymbol oder genau ein Terminalsymbol und eine Variable besitzen.

1.6 Alternative Schreibweise für Grammatiken

Um Grammatiken und deren Regeln einfacher kommunizieren zu können kann man die erzeugten Wörter auch als "Formel" aufschreiben. Dazu folgende Aufgabe.

Aufgabe: Sei $\Sigma = \{0, 1\}$. Leiten Sie mit der folgenden Grammatik das Wort 0101001010 ab.

$S \rightarrow 0T0$
$T \rightarrow 1S1|\epsilon$

Ableitung:

$S \Rightarrow 0T0$
$\Rightarrow 01S10$
$\Rightarrow 010T010$
$\Rightarrow 0101S1010$
$\Rightarrow 01010T01010$
$\Rightarrow 0101001010$

Figure 1.6: Zugehörige Ableitung

Aus dieser Grammatik lassen sich folgende mögliche Wörter ableiten:

- 00
- 010010
- 0101001010
- ...

Da die Grammatik rekursiv ist und damit die Wortliste nicht endlich ist, kann man unmöglich alle Wörter auflisten. Um alle Wörter die erzeugt werden können trotzdem aufschrieben zu können, kann die Menge der Wörter auch formelähnlich aufgeschrieben werden.

$\{(01)^n 00 (10)^n | n \geq 0\}$

Der erste Teil (01) ist genau so oft enthalten wie der letzte Teil (10) kann aber auch weggelassen werden. der mittlere Teil (00) ist immer vorhanden und kann auch nicht wiederholt werden. Durch diese vereinfachte Schreibweise ist es sehr leicht möglich weitere Wörter zu erzeugen. Man muss nur einen Wert für n einsetzen und wiederholt die entsprechenden Teile n-mal.

Aufgabe: Sei $\Sigma = \{0, 1\}$. Entwicklen Sie eine Grammatik, die folgende Sprache erzeugt: $L = \{w \in \Sigma^* / |w|\ \textit{ist gerade}\}$

Ansatz: Das das zu erzeugende Wort (w) ein Σ^* enthalten sein soll, kann es also aus allen möglichen Kombinationen aus Σ (0 und 1) bestehen. Als zusätzliche Beschränkung für die Menge aller Wörter soll der Betrag, also die Länge, des Wortes gerade (durch Zwei teilbar) sein.

Aus dem obigen Ansatz ergeben sich folgende Regeln für die Grammatik.

$S \rightarrow PS|P$ (beliebig viele Pärchen aneinanderreihen)
$P \rightarrow ZZ|\epsilon$ (Pärchen von Zahlen aus Ziffern bilden)
$Z \rightarrow 0|1$ (Ziffern)

$S \Rightarrow PS$
$\Rightarrow PP$
$\Rightarrow ZZP$
$\Rightarrow ZZZZ$
$\Rightarrow 0ZZZ$
$\Rightarrow 01ZZ$
$\Rightarrow 010Z$
$\Rightarrow 0100$

Figure 1.7: Mögliche Ableitung für das Wort 0100

1.7 Ableitungsbaum

Aus dieser Ableitung lässt sich ein Ableitungsbaum erstellen. Dieser stellt strukturiert dar, wie ein Wort abgeleitet wird.

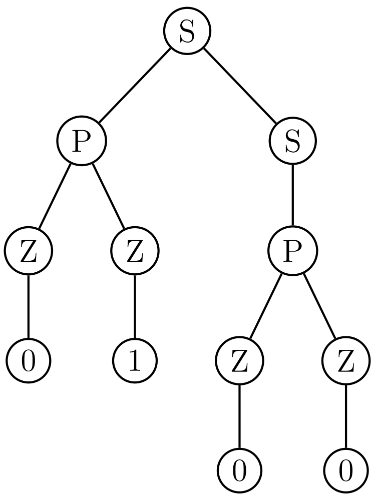

Figure 1.8: Ableitungsbaum für die obige Ableitung

Egal in welcher Reihenfolge abgeleitet wird, bleibt der Ableitungsbaum immer der selbe. Die Ableitung nach $S \Rightarrow PS \Rightarrow ZZS \Rightarrow ...$ oder nach $S \Rightarrow PS \Rightarrow PP \Rightarrow ...$ erzeugen den selben Ableitungsbaum. Anders sieht es aus, wenn eine andere Regelalternative zur Ableitung genutzt wird. Dann kann sich der Ableitungsbaum durchaus unterscheiden.

Die Ableitung $S \Rightarrow PS \Rightarrow PPS \Rightarrow PPP \Rightarrow ZZPP \Rightarrow ZZZZP \Rightarrow ZZZZ\epsilon \Rightarrow ...$ erzeugt den folgenden Ableitungsbaum.

Da es für die hier verwendete Grammatik strukturell unterschiedliche Ableitungs-

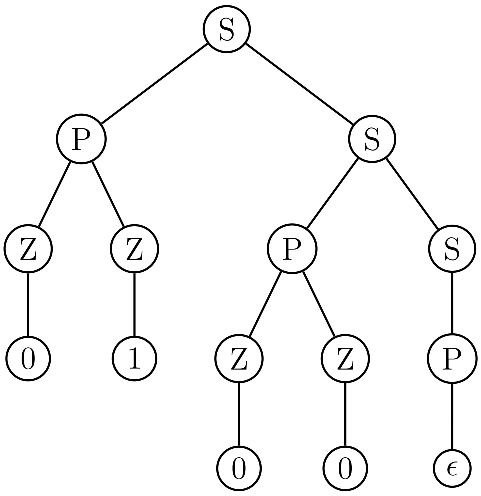

Figure 1.9: Ableitungsbaum für eine alternative Ableitung

bäume gibt, ist die Grammatik mehrdeutig. Daher ist es nicht möglich einen Parser für die erzeugte Sprache zu generieren.

Durch Linksableitung kann Mehrdeutigkeit von Bäumen (Ableitungen) eindeutig bewiesen werden. Die Linksableitung ist eine Konvention, dass von Ableitungsschritt zu Ableitungsschritt die linkeste Variable zuerst abgeleitet wird.

(Bilder der Vorlesung vom 10.04.13 einfügen!)