

Sprachverarbeitung: Übung 15

Syntaktische Analyse mathematischer Ausdrücke

Das Thema dieser Übung ist die syntaktische Analyse einfacher mathematischer Ausdrücke. Die Menge der mathematischen Ausdrücke wird hier als formale Sprache L_M bezeichnet. In der Vorlesung ist gezeigt worden, dass L_M durch eine kontextfreie Grammatik $G_M = (V_N, V_T, P, S)$ beschreibbar ist. Ein Beispiel einer Grammatik für einfache mathematische Ausdrücke ist:

$$\begin{aligned} V_N &= \{S, \text{Exp}, \text{Term}, \text{Fak}, \text{Ident}\} \\ V_T &= V_M \cup V_I = \{+, *, (,)\} \cup \{a, b, c\} \\ P &= \{ S \rightarrow \text{Exp}, \\ &\quad \text{Exp} \rightarrow \text{Term} \mid \text{Term} + \text{Exp}, \\ &\quad \text{Term} \rightarrow \text{Fak} \mid \text{Fak} * \text{Term}, \\ &\quad \text{Fak} \rightarrow \text{Ident} \mid (\text{Exp}), \\ &\quad \text{Ident} \rightarrow a \mid b \mid c \} \end{aligned}$$

Um eine solche (kontextfreie) Grammatik für die Analyse mathematischer Ausdrücke anzuwenden, ist ein Parser erforderlich. Starten Sie den Parser mit dem Matlab-Skript `start_parser_gui.m`.¹ Nach dem Starten erscheint eine grafische Oberfläche, in der Sie in die entsprechenden Felder den Namen der Datei mit der anzuwendenden Grammatik sowie die zu analysierende Symbolsequenz eintragen müssen. Der Ableitungsbaum wird dann im unteren Teil des Fensters dargestellt.

Auf Seite 3 ist eine Beispielgrammatik zu finden, aus der zu sehen ist, wie eine Grammatikdatei aufgebaut wird. Sie können feststellen, dass dieses Beispiel der obigen Grammatik G_M entspricht. Die Grammatik besteht aus zwei Teilen: der Definition der Terminalsymbole und den Produktionsregeln.

Bei der Eingabe der Symbolsequenz sollten Sie beachten, dass der Parser erwartet, dass Terminalsymbole immer durch Leerzeichen getrennt sind. Das ist erforderlich, weil die Symbole auch aus mehreren ASCII-Zeichen bestehen können. Eine korrekte Eingabe ist also zum Beispiel `a + b`.

Aufgabe 1: Testen der gegebenen Grammatik G_M

Testen Sie die Grammatik G_M (ist in der Datei `Gegebenes/ueb15_1.dat` zu finden) mit einigen positiven und negativen Beispielen, also sowohl mit solchen, die zur Sprache $L(G_M)$ gehören, als auch mit nicht zu $L(G_M)$ gehörenden Beispielen.² Verifizieren Sie insbesondere, dass auch mathematische Ausdrücke mit überflüssigen Klammern von dieser Grammatik akzeptiert werden.

¹Mit diesem Matlab-Skript kann der Parser für alle Betriebssysteme gleich gestartet werden. Sonst wird Matlab für diese Übung nicht benötigt.

²Wird ein mathematischer Ausdruck eingegeben, der nicht der Grammatik entspricht, dann gibt der Parser die folgende Meldung aus: *“No valid tree found.”*

Aufgabe 2: Grammatik für mathematische Ausdrücke ohne unnötige Klammern

In der Übung 14 haben Sie die Grammatik G_M aus Aufgabe 1 so verändert, dass die Sprache $L(G'_M)$ der resultierenden Grammatik keine Ausdrücke mit überflüssigen Klammern enthält. Übertragen Sie nun diese Grammatik G'_M in die im Beispiel auf Seite 3 gezeigte Form und testen Sie sie wiederum mit positiven und negativen Beispielen.

Überprüfen Sie auch, ob bei den positiven Beispielen der angezeigte Ableitungsbaum hinsichtlich der Präzedenz (Rangordnung der Operationen $+$ und $*$) korrekt ist. Falls dies nicht der Fall ist, dann müssen Sie Ihre Grammatik korrigieren.

Aufgabe 3: Grammatik für etwas komplexere mathematische Ausdrücke

Schreiben Sie eine Grammatik $G''_M = (V''_N, V''_T, P'', S)$, bei welcher die Menge der Terminalsymbole um das Minuszeichen erweitert ist, also $V''_T = V_T \cup \{-\}$, so dass $L(G''_M)$ auch Ausdrücke wie $a - b$, $-a$, $+a$, $-a - b$, $-a * b$, $a * (-b)$, $-(+a)$, $-(a + b) * (-c)$ usw. umfasst. Nicht enthalten soll $L(G''_M)$ jedoch Ausdrücke wie $a - +b$, $a * -b$ etc. In den Wörtern der Sprache $L(G''_M)$ dürfen folglich nicht zwei Operationszeichen direkt nacheinander stehen, überflüssige Klammern sind jedoch erlaubt.

Testen Sie Ihre Grammatik wiederum mit geeigneten Beispielen und achten Sie insbesondere auch darauf, dass bei den positiven Beispielen der Ableitungsbaum die Präzedenz richtig wiedergibt.

Grammatik G_M zu Aufgabe 1:

```
! Übung 15: Aufgabe 1
!
! Grammatik für mathematische Ausdrücke
!-----
! Erlaubte Operationen:  + * Klammern
!
! Testen dieser Grammatik (Terminalsymbole durch Leerzeichen trennen!)
! a) positive Beispiele:  a + b * ( a + c )
!                          ( a * b ) + c
!                          ( a + b )
! b) negative Beispiele:  a + b ( a + c )
!                          ( b + )

! Terminalsymbole
:LEXICON lexikon

IDENT    "a"
IDENT    "b"
IDENT    "c"
KL1      "("
KL2      ")"
MULT     "*"
PLUS     "+"

:END

! Produktionsregeln
:GRAMMAR grammatik

S ==> EXP *

EXP ==> TERM *
EXP ==> TERM PLUS EXP *

TERM ==> FAK *
TERM ==> FAK MULT TERM *

FAK ==> IDENT *
FAK ==> KL1 EXP KL2 *

:END
```