

Einführung in die Computerlinguistik

Hausaufgabe zu symbolischem Parsing

Abgabe 05.07.2022, 8.30 Uhr

Laura Kallmeyer

Sommer 2022, Heinrich-Heine-Universität Düsseldorf

Aufgabe 1 Betrachten Sie folgende CFG:

$G = \langle \{S, B\}, \{a, b\}, \{S \rightarrow aB \mid bB, B \rightarrow bB \mid b\}, S \rangle$

1. Erlaubt diese Grammatik ein deterministisches Top-Down Parsing? Begründen Sie Ihre Antwort.
2. Geben Sie für die Eingabe $w = ab$ alle Paare von Resteingabe und Stack (ohne Analysestack), die sich beim Top-Down Parsing ergeben. Geben Sie dabei für jedes Paar außerdem an, aus welchem anderen Paar und mit welcher Operation es entstanden ist, ähnlich wie in dem Beispiel auf Folie 6.
Gehen Sie davon aus, dass jede Möglichkeit verfolgt wird, bis man nicht mehr weiterkommt.
Nehmen Sie außerdem an, dass nur Kombinationen erzeugt werden, bei denen der Stack nicht länger als die Resteingabe ist.
3. Woran erkennt der Parser, dass $w = ab$ zu der von der Grammatik generierten Sprache gehört?

Lösung:

1. Nein, da es sowohl zu S als auch zu B mehrere Produktionen gibt, also beim Predict mit S bzw. B auf dem Stack jeweils zwei Möglichkeiten vorliegen.
2. Produktionen: $S \rightarrow aB \mid bB, B \rightarrow bB \mid b$

	Resteing.	Stack	Operation
1.	ab	S	Initialisierung
2.	ab	aB	predict 1.
3.	ab	bB	predict 1.
4.	b	B	scan 2.
5.	b	b	predict 4.
6.	–	–	scan 5.
3. ab ist in der Sprache, da der Parser eine Konfiguration mit leerer Resteingabe und leerem Stack (Konfiguration 6) herleiten konnte.

Aufgabe 2 Betrachten Sie nochmals die Grammatik aus Aufgabe 1.

Wie sieht die Folge von Paaren von Stack und Resteingabe aus, die sich für $w = bbb$ mit einem Shift-Reduce Parser ergibt? Geben Sie auch hier an, welche Operation zu dem jeweiligen Paar geführt hat und aus welchem anderen Paar es entstanden ist.

Es sollen in jeder Konfiguration sämtliche reduce Möglichkeiten durchgeführt werden und immer auch shift, egal ob diese Wege zum erfolgreichen Parse führen oder nicht.

Am besten, Sie machen immer im Wechsel 1. alle möglichen reduce Schritte für die Items mit gleicher Resteingabe, und dann 2. shift für alle Items, die sich hier ergeben haben, was zu einer Gruppe mit neuer Resteingabe führt, auf die dann wieder 1. angewendet werden kann usw.

Die Tabelle sieht dann wie folgt aus:

<i>id</i>	<i>Stack</i>	<i>Resteing.</i>	<i>Operation</i>
1		bbb	
2	b	bb	shift 1
3	B	bb	reduce 2, $B \rightarrow b$
4	bb	b	shift 2
5	Bb	b	shift 3

Lösung:

Produktionen: $S \rightarrow aB \mid bB, B \rightarrow bB \mid b$

id	Stack	Resteing.	Operation
1		bbb	
2	b	bb	shift 1
3	B	bb	reduce 2, $B \rightarrow b$
4	bb	b	shift 2
5	Bb	b	shift 3
6	bB	b	reduce 4, $B \rightarrow b$
7	BB	b	reduce 5, $B \rightarrow b$
8	S	b	reduce 6, $S \rightarrow bB$
8'	B	b	reduce 6, $B \rightarrow bB$
9	bbb	-	shift 4
10	Bbb	-	shift 5
11	bBb	-	shift 6
12	BBb	-	shift 7
13	Sb	-	shift 8
13'	Bb	-	shift 8'
14	bbB	-	reduce 9, $B \rightarrow b$
15	BbB	-	reduce 10, $B \rightarrow b$
16	bBB	-	reduce 11, $B \rightarrow b$
17	BBB	-	reduce 12, $B \rightarrow b$
18	SB	-	reduce 13, $B \rightarrow b$
19	bB	-	reduce 14, $B \rightarrow bB$
20	BB	-	reduce 13, $B \rightarrow b$ oder reduce 15, $B \rightarrow bB$
21	bS	-	reduce 14, $S \rightarrow bB$
22	BS	-	reduce 15, $S \rightarrow bB$
23	S	-	reduce 19, $S \rightarrow bB$
24	B	-	reduce 19, $B \rightarrow bB$

(Die etwas komische Nummerierung liegt daran, dass ich zunächst ein Item (8') vergessen hatte. Es sind also insgesamt 26 verschiedene, eines davon entsteht zweimal.)

Item 24 ist nicht mehr nötig, da eine erfolgreiche Konfiguration vorher schon gefunden wurde. Es ist also auch in Ordnung, wenn 24 nicht mehr gelistet wird.

Aufgabe 3 Betrachten Sie nun folgende CFG:

$$G = \langle \{S, B\}, \{a, b\}, \{S \rightarrow BS \mid SB \mid a \mid b, B \rightarrow AB \mid b, A \rightarrow a\}, S \rangle$$

1. Wie sieht die Chart aus, die sich bei einem CYK-Parsing der Eingabe $w = abab$ mit dieser CFG ergibt? (Nur Recognition.)
2. Gehört das Wort zur Sprache? Begründen Sie Ihre Antwort anhand der Chart.

Lösung:

l					
4	S				
3	S	S			
2	B, S	S	S, B		
1	S, A	S, B	$S, A,$	S, B	
	1	2	3	4	i

1.

2. Ja, da es im Feld mit $i = 1, l = 4$ ein S gibt.