

Parsing

Homework 9 (Tomita), due 21 June 2021

Laura Kallmeyer

SS 2021, Heinrich-Heine-Universität Düsseldorf

Question 1 (Tomita)

The following table is the LR(1) parse table for the CFG with non-terminals $\{A, B, T\}$, terminals $\{a, b\}$, start symbol S and productions 1. $S \rightarrow ABA$, 2. $S \rightarrow aTa$, 3. $T \rightarrow bTb$, 4. $T \rightarrow \epsilon$, 5. $A \rightarrow a$, 6. $A \rightarrow aA$, 7. $B \rightarrow bb$

(The table has multiple entries for some of the fields.)

	a	b	$\$$	S	A	B	T
0	$s1$			4	5		
1	$s8, r4$	$s2, r5$			16		9
2		$s3, r4$					10
3		$s3, r4$					11
4			acc				
5		$s13$				6	
6	$s14$				7		
7			$r1$				
8	$s8$	$r5$			16		
9	$s17$						
10		$s18$					
11		$s19$					
12	$r7$						
13		$s12$					
14	$s14$		$r5$		15		
15			$r6$				
16		$r6$					
17			$r2$				
18	$r3$						
19		$r3$					

Give the trace of the Tomita-parse for $abba$ (with all intermediate stack graphs and all analyses).

Solution:

Stack	analysis
0 s1	
0 — 1 — 1 s2,r5	1: a
0 — 1 — 1 s2	
2 — 5 s13	2: A(1)
0 — 1 — 1 — 3 — 2 s3,r4	
2 — 5 — 3 — 13 s12	3: b
4 — 10 s18	
0 — 1 — 1 — 3 — 2 s3	
2 — 5 — 3 — 13 s12	4: T(ϵ)
4 — 10 — 5 — 18 r3	
0 — 1 — 1 — 3 — 2 — 5 — 3 -	
2 — 5 — 3 — 13 — 5 — 12 r7	5: b
0 — 1 — 1 — 6 — 9 s17	
2 — 5 — 3 — 13 — 5 — 12 r7	6: T(3,4,5)
0 — 1 — 1 — 6 — 9 s17	
2 — 5 — 7 — 6 s14	7: B(3,5)
0 — 1 — 1 — 6 — 9 — 8 — 17 r2	
2 — 5 — 7 — 6 — 8 — 14 r5	8: a
0 — 9 — 4 acc	
2 — 5 — 7 — 6 — 8 — 14 r5	9: S(1,6,8)
0 — 9 — 4 acc	
2 — 5 — 7 — 6 — 10 — 7 r1	10: A(8)
0 — 9 — 4 acc	
11	11: S(2,7,10)
0 — 12 — 4 acc	12: [11,9]

Question 2 (Tomita and loops)

For CYK, we have seen that the chart-based structure sharing we have there means that loops (i.e., derivations $A \xrightarrow{+} A$ for some non-terminal A) are no longer a problem, in contrast to shift-reduce parsing.

Tomita provides some structure sharing. What about loops? Are they a problem or not?

In order to answer this, consider the CFG $\langle \{S, A\}, \{a\}, P, S \rangle$ with productions

$$1.S \rightarrow A \quad 2.S \rightarrow a \quad 3.A \rightarrow S$$

The LR(1) parse table is as follows:

	a	$\$$	A	S
0	$s2$		3	1
1		$r3$	acc	
2		$r2$		
3		$r1$		

1. What would the trace of the Tomita-parse for the word $w = a$ look like? Give only the first 6 stack graphs together with the analyses that get constructed, using the compact parse forest representation.
2. Are loops a problem for this algorithm or not? Justify your answer.

Solution:

	Stack	analysis
	$0 \text{ --- } s2$	
	$0 \text{ --- } \boxed{1} \text{ --- } 2 \quad r2$	$\boxed{1}: a$
	$0 \text{ --- } \boxed{2} \text{ --- } 1 \quad r3, acc$	$\boxed{2}: S(\boxed{1})$
1.	$0 \text{ --- } \boxed{2} \text{ --- } 1 \quad acc$	
	$\quad \quad \boxed{3} \text{ --- } 3 \quad r1$	$\boxed{3}: A(\boxed{2})$
	$0 \text{ --- } \boxed{2} \text{ --- } 1 \quad acc$	
	$\quad \quad \boxed{4} \text{ --- } 1 \quad r3 \text{ acc}$	$\boxed{4}: S(\boxed{3})$
	$0 \text{ --- } \boxed{5} \text{ --- } 1 \quad r3 \text{ acc}$	$\boxed{5}: [\boxed{2}, \boxed{4}]$

2. Yes, they are a problem since, concerning the analysis, entire trees are created. This means that we would have larger and larger trees concerning height. On the graph-based stack, correspondingly, every loop would lead to a new sequence of reduce steps, always with the same productions but leading to yet another pointer, i.e., yet another stack.