Parsing

Homework 3 (Deduction-based Parsing), due 10 May 2021

Laura Kallmeyer

SS 2021, Heinrich-Heine-Universität Düsseldorf

Question 1 (Unger with deduction rules)

Consider a CFG with the following productions: $S \to bSd \mid aT \mid ab, T \to cT \mid b$.

Consider the input $w = ab$ *and the deduction rules for non-directional top-down parsing (= Unger parsing) from slides 12 and 13.*

1. *Give all items the parser generates for this input. For every item, indicate the rule that was used to deduce this item and indicate the antecedent items of this rule.*
2. *How does the parser know whether* $w = ab$ *is in the language generated by the grammar?*

Solution:

1.

id	item	operation	antecedent items
1	$[\bullet S, 0, 2]$	axiom	–
2	$[\bullet a, 0, 1]$	predict	1
3	$[\bullet T, 1, 2]$	predict	1
4	$[\bullet b, 1, 2]$	predict	1
5	$[a\bullet, 0, 1]$	scan	2
6	$[b\bullet, 1, 2]$	scan	4
7	$[T\bullet, 1, 2]$	complete	3,6
8	$[S\bullet, 0, 2]$	complete	1,5,6 or 1,5,7

2. There is a goal item $[S\bullet, 0, 2]$ in the chart, therefore the word is in the language.

Question 2 (Unger deduction rules for GNF)

Consider the following deduction rules, for Unger Parsing of CFGs in Greibach Normal Form:

Axiom: $$\frac{}{[\bullet S, 0, n]} \quad |w| = n$$

Goal item: $[S\bullet, 0, n]$ *where* $n = |w|$

Scan_predict: $$\frac{[\bullet A, i_0 - 1, j]}{[a\bullet, i_0 - 1, i_0], [\bullet A_1, i_0, i_1], \dots, [\bullet A_k, i_{k-1}, i_k]} \quad \begin{array}{l} A \to aA_1 \dots A_k \in P \\ w_{i_0} = a, j = i_k, i_m < i_{m+1} \text{ for } 0 \leq m < k \end{array}$$

Complete: $$\frac{[a\bullet, i_0 - 1, i_0], [\bullet A, i_0 - 1, i_k], [A_1\bullet, i_0, i_1], \dots, [A_k\bullet, i_{k-1}, i_k]}{[A\bullet, i_0 - 1, i_k]} \quad A \to aA_1 \dots A_k \in P$$

1. *What is the time complexity of this algorithm in the length of the input string, i.e., of the fixed recognition problem, under the assumption that the righthand sides of rules have a length* ≤ 3*? Explain your answer by estimating the maximal number of different rule applications possible for a given CFG in GNF with this additional constraint.*
2. *Consider a CFG with the following productions (obtained from the one in Question 1 via transformation into GNF):* $S \to bSD \mid aT \mid aB, T \to cT \mid b, B \to b, D \to d$.

   Consider the input $w = ab$ *and the deduction rules for Unger parsing with CFGs in GNF. Give all items the parser generates for this input. For every item, indicate the rule that was used to deduce this item and indicate the antecedent items of this rule.*

Solution:

1. Scan-Predict: maximal $|P| \cdot n \cdot n \cdot n$ different instances that can be applied since each production contains at most two non-terminals in its righthand side, which means that there are maximally three indices involved, each of them taking at most $n - 1 < n$ different values. The same holds for Complete.

   Therefore, we have a total of $\leq 2 \cdot |P|n^3$ different rule applications where c is a constant depending on the grammar. Consequently, the time complexity of the fixed recognition problem is $\mathcal{O}(n^3)$.

2.

id	item	operation	antecedent items
1	$[\bullet S, 0, 2]$	axiom	–
2	$[a\bullet, 0, 1]$	scan-predict	1
3	$[\bullet T, 1, 2]$	scan-predict	1
4	$[\bullet B, 1, 2]$	scan-predict	1
5	$[b\bullet, 1, 2]$	scan-predict	3 or 4
6	$[T\bullet, 1, 2]$	complete	3,5
7	$[B\bullet, 1, 2]$	complete	4,5
8	$[S\bullet, 0, 2]$	complete	1,2,6 or 1,2,7