

23.11.2021

Seminar 9 – LL(1) parser

Ex.: Given the grammar $G = (\{S, A, B, C, D\}, \{+, *, (,), a\}, P, S)$

P : (1) $S \rightarrow BA$

(2) $A \rightarrow +BA$

(3) $A \rightarrow \epsilon$

(4) $B \rightarrow DC$

(5) $C \rightarrow * DC$

(6) $C \rightarrow \epsilon$

(7) $D \rightarrow (S)$

(8) $D \rightarrow a,$

parse the sequence $w = a * (a + a)$ using an LL(1) parser.

I. Compute FIRST & FOLLOW functions

FIRST

	F0	F1	F2	F3 = F2 = FIRST
S	\emptyset	\emptyset	$\{ (, a \}$	$\{ (, a \}$
A	$\{ +, \epsilon \}$	$\{ +, \epsilon \}$	$\{ +, \epsilon \}$	$\{ +, \epsilon \}$
B	\emptyset	$\{ (, a \}$	$\{ (, a \}$	$\{ (, a \}$
C	$\{ *, \epsilon \}$	$\{ *, \epsilon \}$	$\{ *, \epsilon \}$	$\{ *, \epsilon \}$
D	$\{ (, a \}$	$\{ (, a \}$	$\{ (, a \}$	$\{ (, a \}$

$$\{ (, a \} \{ *, \epsilon \} = \{ (*, (, a*, a) \}$$

FOLLOW @B Toncea Ion-Alin

	L0	L1	L2	L3	L4 = L3 = FOLLOW
S	$\{ \epsilon \}$	$\{ \epsilon,) \}$	$\{ \epsilon,) \}$	$\{ \epsilon,) \}$	$\{ \epsilon,) \}$
A	\emptyset	$\{ \epsilon \}$	$\{ \epsilon,) \}$	$\{ \epsilon,) \}$	$\{ \epsilon,) \}$
B	\emptyset	$\{ +, \epsilon \}$	$\{ +, \epsilon,) \}$	$\{ +, \epsilon,) \}$	$\{ +, \epsilon,) \}$
C	\emptyset	\emptyset	$\{ +, \epsilon \}$	$\{ +, \epsilon,) \}$	$\{ +, \epsilon,) \}$
D	\emptyset	$\{ * \}$	$\{ *, +, \epsilon \}$	$\{ *, +, \epsilon,) \}$	$\{ *, +, \epsilon,) \}$

II. LL(1) table @B Toncea Ion-Alin

	+	*	()	a	\$
S			BA, 1		BA, 1	
A	+BA, 2			€, 3		€, 3
B			DC, 4		DC, 4	
C	€, 6	*DC, 5		€, 6		€, 6
D			(S), 7		a, 8	
+	pop					
*		pop				
(pop			
)				pop		
A					pop	
\$						acc

Obs.:

1) All the **empty** cells of the table above are considered to be filled with **error**, by default (accessing such a cell within the analysis means that the sequence is syntactically incorrect).

2) **Duplicated** pairs within a cell (conflicts) indicate that the grammar is **not LL(1)** and the analysis cannot be performed.

Ex.: $A \rightarrow \alpha\gamma \mid \alpha\beta$! not LL(1)

Transformed to

$A \rightarrow \alpha B$

$B \rightarrow \beta \mid \gamma$

III. Parse the input sequence @B Alexandra T.

(a*(a+a)\$, S\$, €) |- (a*(a+a)\$, BA\$, 1) |- (a*(a+a)\$, DCA\$, 14) |- (a*(a+a)\$, aCA\$, 148) |- (*(a+a)\$, CA\$, 148) |- (*(a+a)\$, *DCA\$, 1485) |- ((a+a)\$, DCA\$, 1485) |- ((a+a)\$, (S)CA\$, 14857) |- ((a+a)\$, S)CA\$, 14857) |- ((a+a)\$, BA)CA\$, 148571) |- ((a+a)\$, DCA)CA\$, 1485714) |- ((a+a)\$, aCA)CA\$, 14857148) |- ((a+a)\$, CA)CA\$, 14857148) |- ((a+a)\$, A)CA\$, 148571486) |- ((a+a)\$, +BA)CA\$, 1485714862) |- (a)\$, BA)CA\$, 1485714862) |- (a)\$, DCA)CA\$, 14857148624) |- (a)\$, aCA)CA\$, 148571486248) |- ()\$, CA)CA\$, 148571486248) |- ()\$, A)CA\$, 1485714862486) |- ()\$,)CA\$, 14857148624863) |- (\$, CA\$, 14857148624863) |- (\$, A\$, 148571486248636) |- (\$, \$, 1485714862486363)|- acc