

Syntaktická analýza s kondenzovaným poučením převzatým z [JP].

[JP] 4.1.1. až 4.1.3 včetně

[PJ] 3.1. až 3.4. včetně

## Syntaktická analýza

Vstup

Řetězec terminálních symbolů gramatiky  $G$ ,  
Rozkladová tabulka  $M$  gramatiky  $G$

Výstup

Posloupnost čísel pravidel v levé derivaci vstupního slova nebo hlášení chyby.

Dokud nenastane *Přijetí* nebo *Chyba* prováděj příslušnou *Akci Expanze* nebo *Srovnání*, ve schématu níže, podle toho, jaká je aktuální *Situace*.

<i>Expanze</i>	<i>Situace</i>	První nepřečtený symbol je $a$ . Na vrcholu zásobníku je neterminál $A$ . V rozkladové tabulce $M$ platí $M[A][a] == \beta, i$ .
	<i>Akce</i>	pop( $A$ ), push( $\beta$ ), print( $i$ ). Symbol $a$ ponecháme na vstupu, “nečteme jej”.
<i>Srovnání</i>	<i>Situace</i>	První nepřečtený symbol je $a$ . Na vrcholu zásobníku je také $a$ .
	<i>Akce</i>	pop( $a$ ), Symbol $a$ “přečteme”, tj. posuneme se na vstupu na další symbol.
<i>Přijetí</i>	<i>Situace</i>	Vstup je celý přečten. Zásobník je prázdný.
	<i>Akce</i>	Úspěšný konec analýzy, na výstupu je posloupnost čísel pravidel v levé derivaci vstupního slova.
<i>Chyba</i>	<i>Situace</i>	1. Při pokusu o expanzi není definováno $M[A][a]$ . 2. Při pokusu o srovnání se neterminál $a$ na vrcholu zásobníku neshoduje s prvním nepřečteným symbolem $b$ . 3. Vstup ještě není celý přečten a zásobník je už prázdný. 4. (jen u jednoduché LL(1) gramatiky) Vstup je celý přečten a zásobník je neprázdný.
	<i>Akce</i>	Neúspěšný konec analýzy, hlášení chyby ve vstupu.

Konfigurace algoritmu syntaktické analýzy je trojice

(nepřečtená část vstupního řetězce, obsah zásobníku, posloupnost čísel pravidel na výstupu).

### Jednoduchá LL(1) gramatika splňuje:

1. Pravá strana každého pravidla začíná neterminálním symbolem.
2. Pravé strany každých dvou různých pravidel se stejnou levou stranou začínají různými terminálními symboly.

Rozkladová tabulka M jednoduché LL(1) gramatiky:

Řádky jsou indexovány všemi neterminály, sloupce jsou indexovány všemi terminály dané gramatiky.

$M[A][a] = \alpha, i$ , pokud  $i$ -té pravidlo má tvar  $A \rightarrow \alpha$  ( $A \in N, \alpha \in (N \cup T)^+$ ),  
jinak  $M[A][a]$  zůstává nedefinováno (odpovídá chybě v analyzovaném řetězci).

### Příklad

Jednoduchá LL(1) gramatika  $G$  obsahující právě pravidla 1. – 4. a její rozkladová tabulka.

1.  $S \rightarrow aASc$
2.  $S \rightarrow b$
3.  $A \rightarrow a$
4.  $A \rightarrow cSAb$

M	a	b	c
S	aASc, 1	b, 2	
A	a, 3		cSAb, 4

Posloupnost konfigurací při syntaktické analýze řetězce  $acbabc$  v gramatice  $G$ :

- (  $acbabc$ ,  $S, \varepsilon$  )
- (  $acbabc$ ,  $aASc, 1$  )
- (  $cbabc$ ,  $ASc, 1$  )
- (  $cbabc$ ,  $cSAbASc, 14$  )
- (  $babc$ ,  $SAbASc, 14$  )
- (  $babc$ ,  $bAbSc, 142$  )
- (  $abc$ ,  $AbSc, 142$  )
- (  $abc$ ,  $abSc, 1423$  )
- (  $bbc$ ,  $bSc, 1423$  )
- (  $bc$ ,  $Sc, 1423$  )
- (  $bc$ ,  $bc, 14232$  )
- (  $c$ ,  $c, 14232$  )
- (  $\varepsilon$ ,  $\varepsilon, 14232$  )

### Cvičení

Napište rozkladovou tabulku uvedené gramatiky a proveďte syntaktickou analýzu řetězce  $aaabbc$ .

1.  $S \rightarrow aAB$
2.  $A \rightarrow ab$
3.  $A \rightarrow aAb$
4.  $B \rightarrow cBd$
4.  $B \rightarrow cd$

### Bezkontextová q-gramatika splňuje:

1. Pravá strana každého pravidla je buď  $\varepsilon$  nebo začíná neterminálním symbolem.
2. Neprázdné (= neobsahující  $\varepsilon$ ) pravé strany každých dvou různých pravidel se stejnou levou stranou začínají různými terminálními symboly.
3. Pro každé pravidlo  $A \rightarrow \varepsilon$  platí, že FOLLOW(A) neobsahuje žádné neterminály, kterými začínají pravá strany pravidel s neterminálem A na levé straně.

FOLLOW(A) je množina všech terminálů, které se v procesu generování nějakého slova mohou vyskytnout za těsně neterminálem A.

Rozkladová tabulka M q-gramatiky:

Řádky jsou indexovány všemi neterminály, sloupce jsou indexovány všemi terminály dané gramatiky a navíc symbolem  $\varepsilon$ .

$M[A][a] = \alpha, i$ , pokud  $i$ -té pravidlo má tvar  $A \rightarrow \alpha$  ( $A \in N, \alpha \in (N \cup T)^+$ ),

$M[A][b] = \varepsilon, i$ , pokud  $i$ -té pravidlo má tvar  $A \rightarrow \varepsilon$  a zároveň  $b \in \text{FOLLOW}(A)$ .

Jinak  $M[A][a]$  zůstává nedefinováno (odpovídá chybě v analyzovaném řetězci).

### Příklad

Bezkontextová q-gramatika obsahující právě pravidla 1. – 4. a její rozkladová tabulka.

1.  $S \rightarrow aAS$
2.  $S \rightarrow b$
3.  $A \rightarrow cAS$
4.  $A \rightarrow \varepsilon$

M	a	b	c	$\varepsilon$
S	aAS, 1	b, 2		
A	$\varepsilon, 4$	$\varepsilon, 4$	cAS, 3	

Některé možné případy generování

$S \rightarrow aAS \rightarrow acASS \rightarrow acAbS \rightarrow \dots$

$S \rightarrow aAS \rightarrow acASS \rightarrow acAaAS \rightarrow \dots$

V obou případech lze zvýrazněný neterminál A expandovat na prázdný řetězec pravidlem 4., tedy  $\text{FOLLOW}(A) = \{a, b\}$ .

### Příklad

Bezkontextová q-gramatika obsahující právě pravidla 1. – 4. a její rozkladová tabulka

1.  $S \rightarrow aA$
2.  $S \rightarrow b$
3.  $A \rightarrow cSa$
4.  $A \rightarrow \varepsilon$

M	a	b	c	$\varepsilon$
S	aA, 1	b, 2		
A	$\varepsilon, 4$		cSa, 3	$\varepsilon, 4$

Některé možné případy generování

$S \rightarrow aA \rightarrow a$

$S \rightarrow aA \rightarrow acSa \rightarrow acaAa \rightarrow \dots$

V prvním případě je zvýrazněný neterminál  $A$  expandován na prázdný řetězec pravidlem 4. a dále už v generovaném slově nic není, tedy  $\varepsilon \in \text{FOLLOW}(A)$ . V druhém případě vidíme  $a \in \text{FOLLOW}(A)$ .

Posloupnost konfigurací při syntaktické analýze řetězce  $acaa$  v dané gramatice:

(  $acaa$ ,  $S$ ,  $\varepsilon$  )  
 (  $acaa$ ,  $aA$ , 1 )  
 (  $caa$ ,  $A$ , 1 )  
 (  $caa$ ,  $cSa$ , 13 )  
 (  $aa$ ,  $Sa$ , 13 )  
 (  $aa$ ,  $aAa$ , 131 )  
 (  $a$ ,  $Aa$ , 131 )  
 (  $a$ ,  $a$ , 1314 )  
 (  $\varepsilon$ ,  $\varepsilon$ , 1314 )

### Cvičení

Napište rozkladovou tabulku uvedené gramatiky a proveďte syntaktickou analýzu řetězců  $accb$ ,  $babb$ ,  $bccbdb$ .

1.  $S \rightarrow aAB$
2.  $S \rightarrow bABC$
3.  $A \rightarrow ccC$
4.  $A \rightarrow \varepsilon$
5.  $B \rightarrow dC$
6.  $B \rightarrow \varepsilon$
7.  $C \rightarrow aCb$
8.  $C \rightarrow b$

### LL(1) gramatika splňuje:

Pro každý neterminál  $A$  a pro každá dvě pravidla  $A \rightarrow \alpha$ ,  $A \rightarrow \beta$  platí

1.  $\text{FIRST}(\alpha) \cap \text{FIRST}(\beta) = \emptyset$ .
2. Pokud z  $\alpha$  lze generovat  $\varepsilon$  a z  $\beta$  nelze generovat  $\varepsilon$ , pak  $\text{FOLLOW}(A) \cap \text{FIRST}(\beta) = \emptyset$ .

$\text{FIRST}(\alpha)$  je množina všech terminálů, kterými mohou začínat řetězce vygenerované z  $\alpha$ . Pokud z  $\alpha$  lze generovat  $\varepsilon$ , pak  $\varepsilon \in \text{FIRST}(A)$ .

Rozkladová tabulka  $M$  LL(1) gramatiky:

Řádky jsou indexovány všemi neterminály, sloupce jsou indexovány všemi terminály dané gramatiky a navíc symbolem  $\varepsilon$ .

$M[A][a] = \alpha, i$ , pokud  $i$ -té pravidlo má tvar  $A \rightarrow \alpha$  a zároveň  $a \in (\text{FIRST}(\alpha) - \{\varepsilon\})$ ,  
 $M[A][b] = \varepsilon, i$ , pokud  $i$ -té pravidlo má tvar  $A \rightarrow \alpha$  a zároveň  $b \in \text{FOLLOW}(A)$  &  $\varepsilon \in \text{FIRST}(\alpha)$ .

Jinak  $M[A][a]$  zůstává nedefinováno (odpovídá chybě v analyzovaném řetězci).

### Příklad

LL(1) gramatika obsahující právě pravidla 1. – 8., její rozkladová tabulka a množiny FIRST/FOLLOW využité v její konstrukci.

1.  $E \rightarrow T E'$
2.  $E' \rightarrow +T E'$
3.  $E' \rightarrow \varepsilon$
4.  $T \rightarrow F T'$
5.  $T' \rightarrow *F T'$
6.  $T' \rightarrow \varepsilon$
7.  $F \rightarrow [E]$
8.  $F \rightarrow a$

M	$a$	$+$	$*$	$[$	$]$	$\varepsilon$
E	TE', 1			TE', 1		
E'		+TE', 2			$\varepsilon$ , 3	$\varepsilon$ , 3
T	FT', 4			FT', 4		
T'		$\varepsilon$ , 6	*FT', 5		$\varepsilon$ , 6	$\varepsilon$ , 6
F	$a$ , 8			[E], 7		

FIRST(TE') = {a, [}  
FIRST(+TE') = {+}  
FIRST(FT') = {a, [}  
FIRST(\*FT') = {\*}  
FOLLOW(E') = { $\varepsilon$ , ]}  
FIRST(E) = {[}  
FOLLOW(T') = { $\varepsilon$ , ], +}  
FIRST(a) = {a}

Posloupnost konfigurací při syntaktické analýze řetězce  $a+a$  v dané gramatice:

- (  $a+a$ , E,  $\varepsilon$  )
- (  $a+a$ , T'E', 1 )
- (  $a+a$ , FT'E', 14 )
- (  $a+a$ , aT'E', 148 )
- (  $+a$ , T'E', 148 )
- (  $+a$ , E', 1486 )
- (  $+a$ , +TE', 14862 )
- (  $a$ , TE', 14862 )
- (  $a$ , FT'E', 148624 )
- (  $a$ , aT'E', 1486248 )
- (  $\varepsilon$ , T'E', 1486248 )
- (  $\varepsilon$ , E', 14862486 )
- (  $\varepsilon$ ,  $\varepsilon$ , 148624863 )

**Výpočet funkce FIRST( $\alpha$ ), kde  $\alpha = X_1X_2\dots X_n \in (N \cup T)^*$**

(a)  $F := \{ .X_1X_2\dots X_n \}$

(b) Pro prvek F, v němž za tečkou bezprostředně následuje neterminál, přidáme do F všechna pravidla s tímto neterminálem vlevo a před jejich pravou stranu vložíme tečku.

(c) Označme B levou stranu takového pravidla v F, kde je tečka zapsána na konci. Vezmeme všechny prvky F, v nichž je tečka před B, vytvoříme nové prvky tak, že tečku v každém posuneme za B a nové prvky vložíme do F.

(d) Opakujeme (b) a (c) dokud lze do F přidávat další prvky.

FIRST( $\alpha$ ) potom obsahuje

1. Všechny terminální symboly, které se vyskytnou bezprostředně za nějakou tečkou v F.
2. Prázdný řetězec  $\varepsilon$ , pokud je v F obsažena posloupnost  $\alpha$  s tečkou na konci.

**Výpočet funkce FOLLOW(A), kde A je neterminál.**

(o) Vytvoříme množinu  $N_\varepsilon$  všech neterminálů, z nichž lze generovat  $\varepsilon$ .

(a) Vytvoříme fiktivní pravidlo  $A \rightarrow A$  a položíme  $F := \{ A \rightarrow A. \}$

(b) Pro prvek F s levou stranou B a s tečkou na konci pravé strany vezmeme všechna pravidla gramatiky, v nichž se B vyskytuje napravo, a vložíme je do F s tečkou zapsanou bezprostředně za B.

(c) Pro prvek F obsahující bezprostředně za tečkou neterminál B vezmeme všechna pravidla gramatiky, v nichž se B vyskytuje nalevo, a vložíme je do F s tečkou zapsanou před pravou stranou.

(d) Pro prvek F obsahující bezprostředně za tečkou neterminál B z množiny  $N_\varepsilon$  přidáme do F další prvek vzniklý posunutím tečky za B

(e) Opakujeme (b) a (c) a (d) dokud lze do F přidávat další prvky.

FOLLOW(A) potom obsahuje

1. Všechny terminální symboly, které se vyskytnou bezprostředně za nějakou tečkou v F.
2. Prázdný řetězec  $\varepsilon$ , pokud je v F obsaženo pravidlo  $S \rightarrow \alpha$  s tečkou na konci (S je startovní symbol).