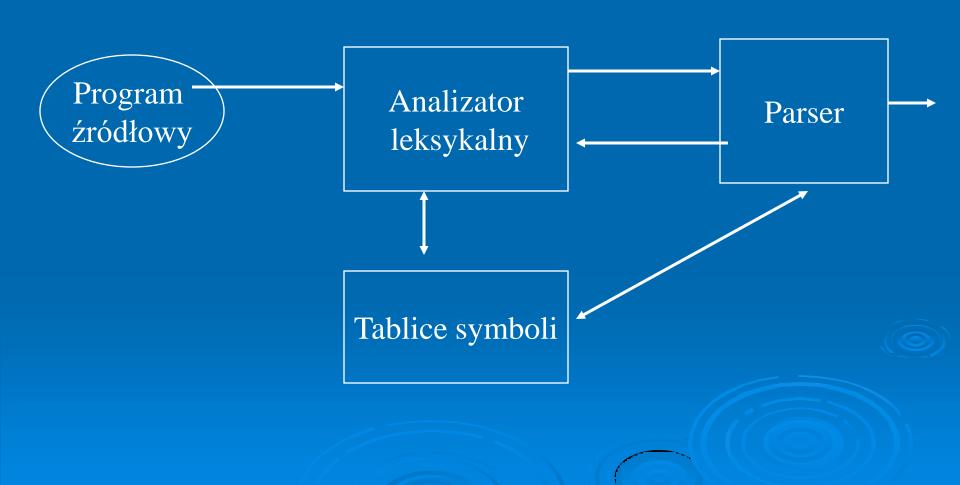
Metody Kompilacji Wykład 8 Analiza Syntaktyczna cd

Wstęp

Parser dostaje na wejściu ciąg tokenów od analizatora leksykalnego i sprawdza: czy ciąg ten może być generowany przez gramatykę.



Wstęp

➤ Istnieją trzy główne rodzaje parserów dla gramatyk: uniwersalny, analiza zstępująca(top-down) oraz analiza wstępująca (bottom-up).

- Wstęp
- Metoda uniwersalna pozwala na zastosowanie dowolnej gramatyki.

Takie ogólne metody jednak mają dużą złożoność obliczeniwą i nie mogą być stosowane do tworzenia kompilatorów przemysłowych.

Wstęp

Metody stosowane do tworzenia kompilatorów przemysłowych są oparte na analizie zstępującej (*top-down*) oraz analizie wstępującej *bottom-up*).

Wyprowadzenia

> Rozważmy następującą gramatykę $E \rightarrow E + E / E * E / -E | (E) | id$

 \triangleright Zamianę symbolu E przez napis -E zapisujemy jako:

$$E \Longrightarrow -E$$

i czytamy: "E wyprowadza -E."

- Wyprowadzenia
- Inne przykłady wyprowadzeń:

$$E * E \Rightarrow (E) * E$$
 lub
 $E * E \Rightarrow E * (E)$

Wyprowadzenia

Możemy wziąć pojedynczy symbol *E* i wielokrotnie stosować produkcje w dowolnej kolejności.

Na przykład,

$$E \Rightarrow -E \Rightarrow -(E) \Rightarrow -(id)$$

Wyprowadzenia

Załóżmy, że nieterminal A znajduje się w środku napisu:

$$\alpha A \beta$$
,

gdzie α and β są to dowolne podnapisy.

Przypuśćmy, że $A \rightarrow \gamma$ jest to produkcja, wtedy wyprowadzenie ma postać:

$$\alpha A\beta \Rightarrow \alpha \gamma \beta$$
.

➤ Symbol ⇒ oznacza wyprowadzenie bezpośrednie.

- Wyprowadzenia
- > Gdy mamy sekwencję wyprowadzeń:

$$a_1 \Rightarrow a_2 \Rightarrow \dots \Rightarrow a_n$$

mówimy, że a_1 wyprowadza a_n co można zapisać również jako

$$a_1 \Rightarrow^* a_n$$

gdzie symbol ⇒* oznacza "zero lub więcej wyprowadzeń".

- Wyprowadzenia
- ▶ Będziemy stosować również symbol ⇒+, który oznacza "co najmniej jedno lub więcej wyprowadzeń".

Wyprowadzenia

▶ Jeśli $S \Rightarrow +a$, gdzie S jest to symbol startowy w gramatyce G, to mówimy, że a jest to forma zdaniowa gramatyki G.

Uwaga: forma zdaniowa może zawierać jednocześnie symbole terminalne jak i nieterminalne.

Wyprowadzenia

> Zdaniem gramatyki *G* jest forma zdaniowa, której wszystkie symbole są terminalami.

➤ Język, generowany przez gramatykę *G*, jest to zbór wszystkich zdań generowanych przez *G*.

Rodzaje Wyprowadzeń

- 1. W wyprowadzeniach lewostronnych(*leftmost derivations*), zawsze wybieramy do zamiany pierwszy nieterminal z lewej strony formy zdaniowej.
- ► Jeśli $\alpha \Rightarrow \beta$ w kroku, w którym zamieniany jest skrajnie lewy nieterminal z α , to wyprowadzenie takie zapisujemy: $\alpha \Rightarrow_{lm} \beta$.

Rodzaje Wyprowadzeń

2. W wyprowadzeniach prawostronnych (*right most derivations*), zawsze wybieramy do zamiany pierwszy nieterminal z prawej strony formy zdaniowej.

Jeśli $\alpha \Rightarrow \beta$ w kroku, w którym zamieniany jest skrajnie prawy nieterminal z α , to wyprowadzenie takie zapisujemy: $\alpha \Rightarrow_{rm} \beta$.

Zbiór First

Tbiór FIRST(α), gdzie α jest dowolnym ciągiem symboli gramatycznych, jest to zbiór terminali, od których zaczynają się ciągi wyprowadzane z α .

 \triangleright Jeśli $\alpha \Rightarrow^* \varepsilon$, to ε należy do FIRST(α).

```
1.exp \rightarrow term \ exp
2.exp' \rightarrow addop \ term \ exp' \ \varepsilon
3.addop \rightarrow + | -
4.term \rightarrow factor\ term'
5.term' \rightarrow mulop factor term' \varepsilon
6.mulop \rightarrow *
7.factor \rightarrow (exp) \mid num
exp \xrightarrow{} term exp' \rightarrow term \rightarrow
                                              Numery
factor term, 5/2 factor \rightarrow (exp)
                                              produkcji
```

		First	
	exp	$\pi(,\dots)$	
_	exp'/	$\longrightarrow \mathcal{E},.$	
	gádop	+ ,-,	
/	term	(, num,	••
	term'	E,	
	mulop	*	
	factor	(,	
		num,	_

```
1.exp \rightarrow term \ exp'
```

$$2.exp' \rightarrow addop \ term \ exp' \mid \varepsilon$$

$$3.addop \rightarrow + \bot - \underline{}$$

 $4.term \rightarrow factor\ term'$

5.term' \rightarrow mulop factor term' | ε

 $6.mulop \rightarrow *$

 $7.factor \rightarrow (exp) \mid num$

	First	
exp	(,	
exp'	E,	
addop	>+ >-,	
term	(, num,	••
term'	<i>E,</i>	
mulop	*	
factor	(,	
	num,	

```
1.exp \rightarrow term \ exp
2.exp' \rightarrow addop \ term \ exp' \mid \varepsilon
3.addop \rightarrow + | -
4.term \rightarrow factor\ term'
5.term' \rightarrow mulop factor term' \mid \varepsilon
6.mulop \rightarrow *
7.factor \rightarrow (exp) \mid num
    term \rightarrow factor term' \rightarrow
    factor \rightarrow (exp) / num
```

	First	
exp	(,	
exp'	<i>E</i> ,	
addop	+ ,-,	
term	→ (ೄ num,	••
term	E,	
mulop	> *	
factor	(,	
	num,	١.

- $1.exp \rightarrow term \ exp'$
- $2.exp' \rightarrow addop \ term \ exp' \mid \varepsilon$
- $3.addop \rightarrow + | -$
- $4.term \rightarrow factor\ term'$
- 5.term' \rightarrow mulop factor term' | ε
- $6.mulop \rightarrow *$
- $7.factor \rightarrow (exp) \mid num$

	First	
exp	(,	
exp'	E,	
addop	+ ,-,	
term	(, num,	••
term'	<i>E,</i>	
mulop	*	
factor	→ (,	
	>num,	

Analiza Syntaktyczna Zbiór FOLLOW

➤ Zbiór FOLLOW(A) dla nieterminala A, jest to zbiór terminali a, które mogą wystąpić bezpośrednio na prawo od A w pewnej formie zdaniowej.

Analiza Syntaktyczna Zbiór FOLLOW

➤ Ponadto, jeśli *A* jest skrajnym symbolem w formie zdaniowej, to dodajemy symbol \$ do zbioru FOLLOW(*A*),

gdzie \$ jest to symbol specjalny oznaczający koniec napisu ("*endmarker*") i nie należący do żadnej gramatyki.

Analiza Syntaktyczna Zbiór FOLLOW

➤ Ponadto, jeśli A jest symbolem startowym, to dodajemy symbol \$ do zbioru FOLLOW(A), gdzie \$ jest to symbol specjalny oznaczający koniec napisu ("endmarker,,) i nie należący do żadnej gramatyki.

expr jest symb. startowym

 $\overline{exp} \rightarrow term \ exp'$ $exp' \rightarrow addop \ term \ exp' \mid \varepsilon$ $addop \rightarrow + |$ $term \rightarrow factor term'$ $term' \rightarrow mulop factor term'$ $mulop \rightarrow *$ $factor \rightarrow (exp) \mid num$

	First	Follow
exp	(num	\\\ \ \
exp'	$\boxed{\varepsilon}$ + -	\$)
addop	+ -	
term	(num	+ - \$)
term'	E *	
mulop	*	
factor	(num	

expr' jest symb. skrajnym

 $exp \rightarrow term \ exp'$ $exp' \rightarrow addop \ term \ exp' \ \mathcal{E}$ $addop \rightarrow + |$ $term \rightarrow factor term'$ term' → mulop factor term' $mulop \rightarrow *$ $factor \rightarrow (exp) \mid num$

	First	Follow
exp	(num	\$)
exp'	ε + -	<u>-\$</u>
addop	+ -	
term	(num	+ - \$)
term'	£ *	
mulop	*	
factor	(num	

 $factor \rightarrow (exp) \rightarrow (term exp')'$

```
exp \rightarrow term \ exp'
exp' \rightarrow addop \ term \ exp' \mid \varepsilon
addop \rightarrow + | -
term \rightarrow factor term'
term' → mulop factor term'
mulop \rightarrow *
factor \rightarrow (exp) \mid num
```

	First	Follow
exp	(num	\$)
exp'	E + -	\$)
addop	+ -	
term	(num	+ - \$)
term'	£*	
mulop	*	
factor	(num	

```
exp \rightarrow term \ exp' \rightarrow term \ addop \ term \ exp' \rightarrow term \ +(-) \ term \ exp'
```

+(-) oznacza + lub -

```
exp \rightarrow term exp'
exp' \rightarrow addop \ term \ exp' \mid \varepsilon
addop \rightarrow + | -
term \rightarrow factor term'
term' → mulop factor term'
mulop \rightarrow *
factor \rightarrow (exp) \mid num
```

	First	Follow
exp	(num	\$)
exp'	ε + -	\$)
addop	+ -	
term	(num	+ - ,\$)
term'	E *	
mulop	*	
factor	(num	

 $exp \rightarrow term \ exp' \rightarrow term \ exp'$

term jest symb. skrajnym

```
exp \rightarrow term exp'
exp' \rightarrow addop \ term \ exp' \mid \varepsilon
addop \rightarrow + | -
term \rightarrow factor term'
term' → mulop factor term'
mulop \rightarrow *
factor \rightarrow (exp) \mid num
```

	First	Follow
exp	(num	\$)
exp'	ε + -	\$)
addop	+ -	
term	(num	+ - \$)
term'	E *	
mulop	*	
factor	(num	

 $factor \rightarrow (exp) \rightarrow (term exp') \rightarrow (term)$

Analiza Syntaktyczna Gramatyki LL(1)

➤ Pierwsze "L" w LL (1) oznacza skanowanie wejścia od lewej do prawej strony, drugie "L" oznacza zastosowanie wyprowadzenia lewostronnego, i "1" oznacza, że w każdym kroku parsowania tylko jeden symbol wejściowy jest brany pod uwagę.

Gramatyki LL(1)

Klasa gramatyk LL(1) jest dość bogata, aby uwzględnić większość konstrukcji programistycznych.

Ale są również ograniczenia: gramatyki rekurencyjne lewostronnie oraz gramatyki niejednoznaczne nie należą do gramatyk LL(1).

Analiza Syntaktyczna Gramatyki LL(1)

- ➤ Gramatyka *G* należy do LL(1) wtedy i tylko wtedy gdy dla dwóch różnych produkcji
 - $A \rightarrow \alpha/\beta$, należących do G, są spełnione następujące warunki:
 - 1. Dla każdego terminala \mathbf{a} , z α i β nie daje się jednocześnie wyprowadzić ciągu rozpoczynającego się od \mathbf{a} .

Gramatyki LL(1)

- \triangleright 2. Co najwyżej z jednego z α i β daje się wyprowadzić napis pusty.
 - 3. Jeśli $\beta \Rightarrow *\varepsilon$ to α nie wyprowadza napisu, którego pierwszy symbol jest terminalem i który znajduje się w zbiorze FOLLOW(A).

Jeśli $\alpha \Rightarrow \varepsilon$ to β nie wyprowadza napisu, którego pierwszy symbol jest terminalem i który znajduje się w zbiorze FOLLOW(A).

Analiza Syntaktyczna Gramatyki LL(1)

- Pierwsze dwa warunki oznaczają, że zbiory $FIRST(\alpha)$ i $FIRST(\beta)$ są rozłączne.
- Trzeci warunek oznacza, że jeśli ε należy do zbioru FIRST(β), to zbiory FIRST(α) i FOLLOW(A) są rozłączne; podobnie gdy ε należy do zbioru FIRST(α).

Tworzenie tablicy do parsowania przewidującego

Parsowanie z wykorzystaniem gramatyk LL(1) wymaga w pierwszej kolejności utworzenia tablicy parsowania.

Tworzenie tablicy do parsowania przewidującego

Następujące reguły są stosowane do tworzenia tablic parsowania

- 1. Dla każdej produkcji $A \rightarrow \alpha$ z gramatyki wykonaj kroki podane niżej.
- 2. Dla każdego terminala a z FIRST(α), dodaj produkcję $A \rightarrow \alpha$ do macierzy M[A,a]
 - 3. Jeśli ε należy do FIRST(α), to
 dla każdego terminala b z FOLLOW(A),
 dodaj A → α do macierzy M[A,b]

Tworzenie tablicy do parsowania przewidującego

Następujące reguły są stosowane do tworzenia tablic

- 4. Jeśli ε należy do FIRST(α) oraz \$ należy do FOLLOW(A), to dodaj A $\rightarrow \alpha$ do M[A, \$].
- 5. Jeśli jakieś pole tablicy jest puste po wykonaniu kroków 2-4, oznacza to, że jeśli zostanie wybrane takie pole, to analizator zwraca błąd i kończy parsowanie.

Przykład: Tworzenie Tablicy parsowania

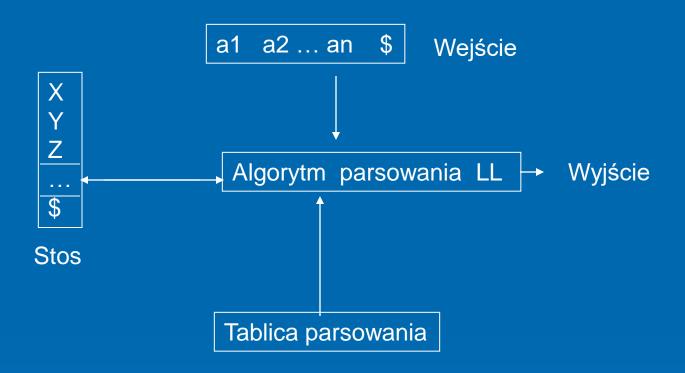
avn	First	Follow
exp,	{(, num}	{\$,)} {\$ \1
exp'	$\{+,-,\ \mathcal{E}\}$	<i>{\$,)}</i>
addop		{(<i>,</i> num}
term	{(, num}	{+,-,),\$}
term'	$\{^*, \ \mathcal{E}\}$	{+,-,),\$}
mulop	{*}	{(<i>,</i> num}
factor	{(, num}	{*,+,-,),\$
	→ term exp' addop term exp &	o'
4 addop -	→ +	
5 addop -	→ -	
-	factor term'	
	mulop factor	term'
8 term' -	ε	
9 mulop -	→ *	
10 factor	$\rightarrow (exp)$	
11 factor	\rightarrow num	

	()	+	-	*	n	\$
ехр	1					1	
exp'		3	2	2			3
addop			4	5			
term	6					6	
term'		8	8	8	7		8
mulop					9		
factor	10					11	

Tworzenie tablicy do parsowania przewidującego

- Dla każdej gramatyki LL(1), każde pole(wpis) w tablicy parsowania posiada tylko jedną produkcję lub jest puste (oznacza błąd).
- ➤ Jeśli dla danej gramatyki jakieś pole tablicy zawiera dwie lub więcej produkcji, to taka gramatyka nie należy do klasy LL(1).

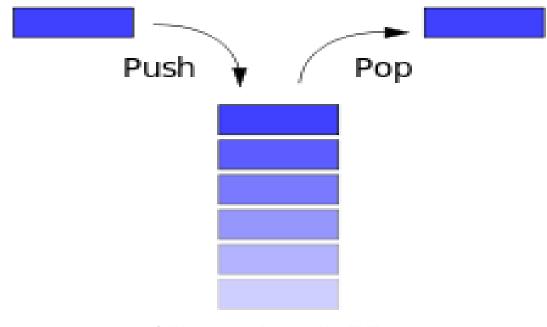
Architektura parsera LL(1)



Analiza Syntaktyczna

Operacje na stosie

- Push -- odłożenie obiektu na stos;
- Pop ściągnięcie obiektu ze stosu i zwrócenie jego wartości;



Algorytm parsowania

• Załóżmy, że:

zmienna o nazwie top zawiera zawartość, która znajduje się na szczycie stosu(pierwsza wartość do odczytu);

zmienna o nazwie input przechowuje bieżący terminal zdania wejściowego;

M[top, input] oznacza zawartość wpisu tablicy parsowania dla argumentów top i input.

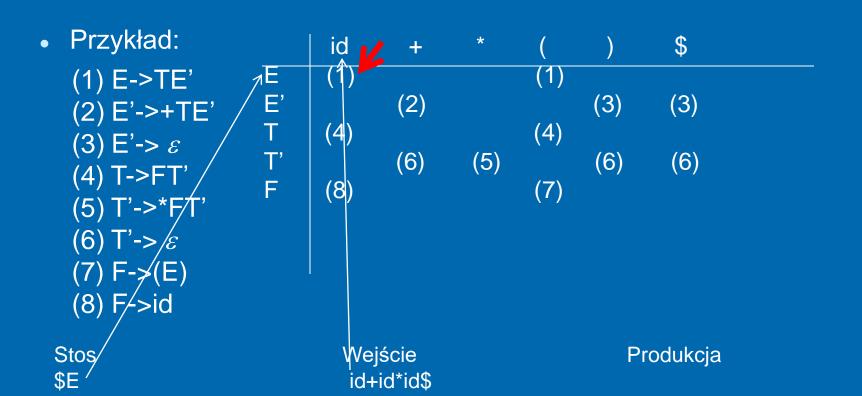
Wtedy algorytm parsowania przedstawia następny slajd.

Algorytm parsowania

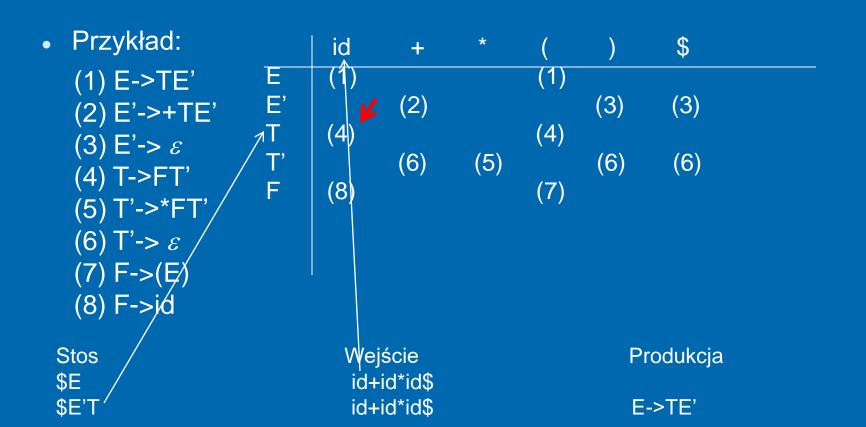
```
odłóż na stos dwa symbole: '$ Symbol Startowy' ($ oznacza koniec napisu).
```

- 1) if top == input == \$ then akceptuj zdanie wejściowe jest poprawne;
- 2) if top == input then
 zdejmij wartość ze szczytu stosu; odczytaj
 następny symbol wejściowy i podstaw go pod
 zmienną input; goto 1;
- 3) If top is nonterminal

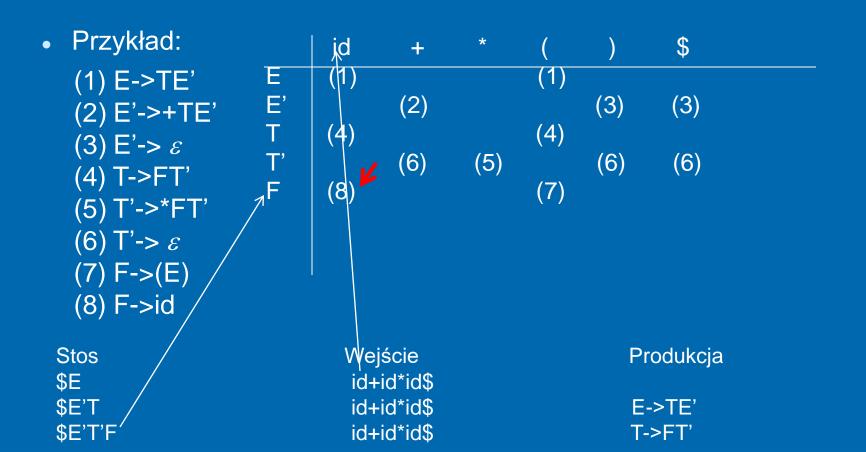
```
if M[top, input] jest to produkcja, to zastąp
wartość na szczycie stosu przez ciało tej
produkcji; goto 1;
else error // zawartość M[top, input] jest polem
pustym
Else error //top zawiera terminal
```



Wyprowadzenie lewostronne: E=>TE'



Wyprowadzenie lewostronne: E=>TE'=>FT'E



Wyprowadzenie lewostronne: E=>TE'=>FT'E'=>idT'E

- Przykład:
 - (1) E->TE'
 - (2) E'->+TE'
 - (3) E'-> ε
 - (4) T->FT'
 - (5) T'->*FT'
 - (6) T'-> ε
 - (7) F -> (E)
 - (8) F->id

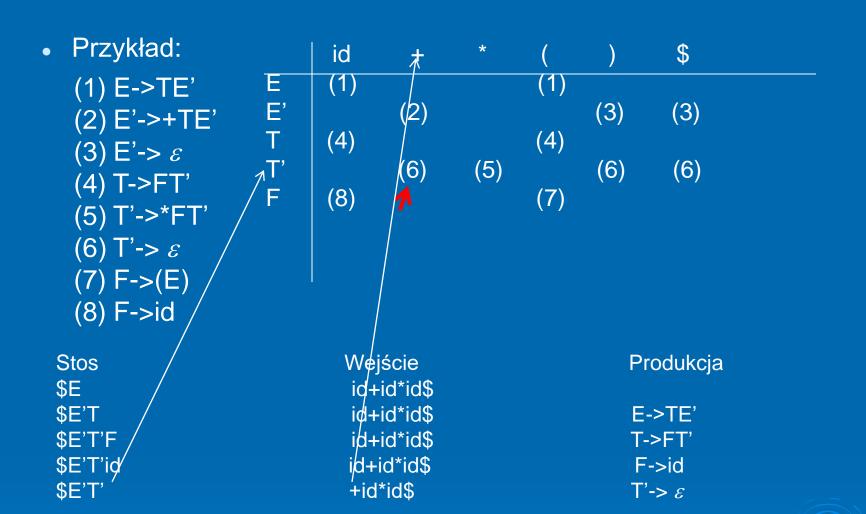
\$ id (1) (1) E T T (2)(3)(3)(4)(4)(6)(5)(6)(6)(8)(7)

Stos \$E \$E'T \$E'T'F \$E'T'id Wejście
id+id*id\$
id+id*id\$
id+id*id\$
id+id*id\$

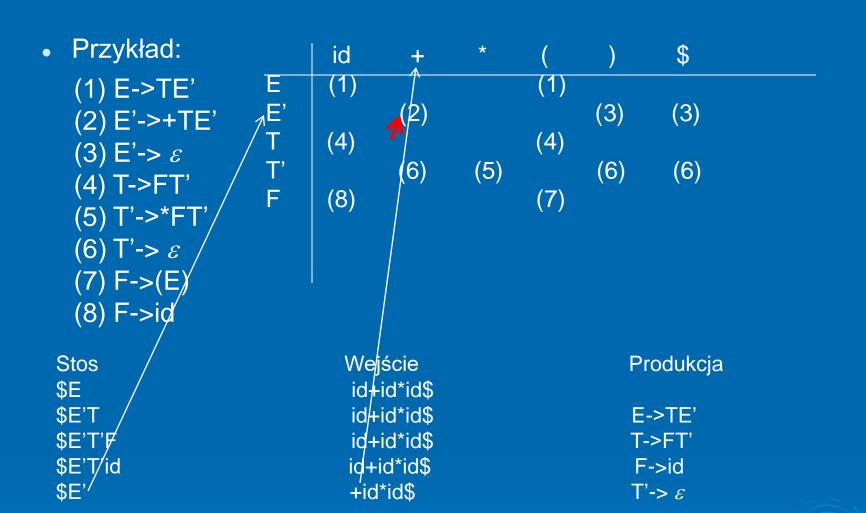
Produkcja

E->TE' T->FT' F->id

Wyprowadzenie lewostronne: E=>TE'=>FT'E'=>idT'E'



Wyprowadzenie lewostronne: E=>TE'=>FT'E'=>idT'E'=>idE'



Wyprowadzenie lewostronne: E=>TE'=>FT'E'=>idT'E'=>idE'

Przykład:

- (1) E->TE'
- (2) E'->+TE'
- (3) E'-> ε
- (4) T->FT'
- (5) T'->*FT'
- (6) T'-> ε
- (7) F -> (E)

id

(8)

(7)

- \$

- (1) (1)
 - (2)

- (3)(3)
- (4)(4)
- E'T' (6)(5)F

(6)(6)

- (8) F->id

Stos

\$E

\$E'T

\$E'T'F

\$E'T'id

\$E'

\$E'T+

Wejście

id+id*id\$

id+id*id\$

id+id*id\$ id+id*id\$

+id*id\$

+id*id\$

Produkcja

E->TE'

T->FT'

F->id

 $T' -> \varepsilon$

E'->+TE'

Wyprowadzenie lewostronne:

E=>TE'=>FT'E'=>idT'E'=>id +TE'

Przykład:

- (1) E->TE' (2) E'->+TE'
 - ,

- id +
- *

(3)

\$

E' T T'

F

(1) (2) (1)

(7)

(3)

- (3) E'-> ε

(4)

(8)

- (6)
- (4) (5)
- (6) (6)

- (4) T->FT'
- (5) T'->*FT'
- (6) T'-> ε
- (7) F -> (E)
- (8) F->id

\$E

\$E'T

Stos

\$E'T'F

\$E'T'id

\$E'

\$E'T

Wejście

id+id*id\$

id+id*id\$

id+id*id\$

id+id*id\$

+id*id\$

+id*id\$

Produkcja

E->TE'

T->FT'

F->id

Τ'-> ε

E'->+TE'

Wyprowadzenie lewostronne:

Dziękuję za uwagę