

# Analiza wstępująca. Gramatyki operatorowe. Gramatyki $LR(k)$ Języki formalne i techniki translacji - Wykład 10

Maciek Gębala

10 grudnia 2019

Maciek Gębala Analiza wstępująca. Gramatyki operatorowe. Gramatyki  $LR(k)$

## Analiza redukująca

Analiza redukująca buduje drzewo wyprowadzenia dla wejścia, zaczynając od liści (od dołu) i przechodząc w górę, aż do korzenia (do góry). W każdym kroku redukcji pewien podciąg pasujący do prawej strony produkcji jest zastępowany symbolem z lewej strony tej produkcji. Jeśli podciągi wybieramy prawidłowo to otrzymujemy prawostronne wyprowadzenie.

- Metoda pierwszeństwa operatorów.
- Analiza LR.

Maciek Gębala Analiza wstępująca. Gramatyki operatorowe. Gramatyki  $LR(k)$

## Przykład

- Weźmy gramatykę

$$\begin{aligned} S &\rightarrow aABe \\ A &\rightarrow Abc|b \\ B &\rightarrow d \end{aligned}$$

- Weźmy słowo  $abbcde$ .
- Pierwszym symbolem który możemy zredukować jest  $b$  i otrzymujemy  $aAbcde$ .
- Teraz możemy zredukować  $Abc$  i otrzymujemy  $aAde$ .
- Teraz  $d$  i  $aABe$ .
- Stąd mamy prawostronne wyprowadzenie

$$S \Rightarrow aABe \Rightarrow aAde \Rightarrow aAbcde \Rightarrow abbcde$$

Maciek Gębala Analiza wstępująca. Gramatyki operatorowe. Gramatyki  $LR(k)$

## Uchwyty

Uchwyt  $\gamma$  to produkcja  $A \rightarrow \beta$  i pozycja w  $\gamma$ , na której znajduje się ciąg symboli  $\beta$ , który po zastąpieniu przez  $A$  da nam ciąg poprzedzający  $\gamma$  w wyprowadzeniu prawostronnym.

- Jeśli  $S \Rightarrow^* \alpha Aw \Rightarrow \alpha \beta w$ , to  $A \rightarrow \beta$  na pozycji po  $\alpha$  jest uchwytem  $\alpha\beta w$ .
- Ciąg symboli po prawej stronie uchwytu zawiera tylko terminale.
- Jeśli gramatyka jest jednoznaczna to w trakcie redukcji mamy zawsze tylko jeden uchwyt.

Maciek Gębala Analiza wstępująca. Gramatyki operatorowe. Gramatyki  $LR(k)$

## Przykład - Shift-Reduce-Parsing

- $E \rightarrow E + E \mid E * E \mid (E) \mid id$
- Słowo  $id_1 + id_2 * id_3$ .
- Wyprowadzenie prawostronne

$$\begin{aligned} E &\Rightarrow E + E \Rightarrow E + E * E \\ &\Rightarrow E + E * id_3 \Rightarrow E + id_2 * id_3 \Rightarrow id_1 + id_2 * id_3 \end{aligned}$$

- Redukcja

| Słowo                | Uchwyt  | Produkcja do redukcji |
|----------------------|---------|-----------------------|
| $id_1 + id_2 * id_3$ | $id_1$  | $E \rightarrow id$    |
| $E + id_2 * id_3$    | $id_2$  | $E \rightarrow id$    |
| $E + E * id_3$       | $id_3$  | $E \rightarrow id$    |
| $E + E * E$          | $E * E$ | $E \rightarrow E * E$ |
| $E + E$              | $E + E$ | $E \rightarrow E + E$ |
| $E$                  |         |                       |

Maciek Gębala

Analiza wstępująca, Gramatyki operatorowe, Gramatyki LR(k)

Notatki

## Parsowanie z użyciem stosu

Stos używamy do pamiętania symboli z gramatyki. Uchwyt jest zawsze na szczycie stosu.

| Stos             | input                   | czynność                             |
|------------------|-------------------------|--------------------------------------|
| \$               | $id_1 + id_2 * id_3 \$$ | shift                                |
| $\$id_1$         | $+ id_2 * id_3 \$$      | redukcja przez $E \rightarrow id$    |
| $\$E$            | $+ id_2 * id_3 \$$      | shift                                |
| $\$E +$          | $id_2 * id_3 \$$        | shift                                |
| $\$E + id_2$     | $* id_3 \$$             | redukcja przez $E \rightarrow id$    |
| $\$E + E$        | $* id_3 \$$             | shift                                |
| $\$E + E *$      | $id_3 \$$               | shift                                |
| $\$E + E * id_3$ | $\$$                    | redukcja przez $E \rightarrow id$    |
| $\$E + E * E$    | $\$$                    | redukcja przez $E \rightarrow E * E$ |
| $\$E + E$        | $\$$                    | redukcja przez $E \rightarrow E + E$ |
| $\$E$            | $\$$                    | Accept                               |

Maciek Gębala

Analiza wstępująca, Gramatyki operatorowe, Gramatyki LR(k)

Notatki

## Problemy w trakcie analizy

- Konflikt przesunięcie/redukcja** – mimo znajomości całego stosu i następnego symbolu wejściowego nie można zdecydować, czy wykonać przesunięcie czy redukcję.
- Konflikt redukcja/redukcja** – możemy wykonać kilka możliwych redukcji.

Maciek Gębala

Analiza wstępująca, Gramatyki operatorowe, Gramatyki LR(k)

Notatki

## Gramatyki operatorowe

Gramatyka jest operatorowa jeśli żadna prawa strona produkcji nie zawiera dwóch sąsiednich nieterminali.

Gramatyka nie-operatorowa

$$\begin{aligned} E &\rightarrow EAE \mid id \\ A &\rightarrow + \mid - \mid * \mid / \end{aligned}$$

Równoważna gramatyka operatorowa

$$E \rightarrow E + E \mid E - E \mid E * E \mid E / E \mid id$$

Maciek Gębala

Analiza wstępująca, Gramatyki operatorowe, Gramatyki LR(k)

Notatki

## Priorytety operatorów

Między pewnymi parami terminali definiujemy trzy rozłączne relacje

- $a < b$  a ma mniejszy priorytet niż b
- $a \doteq b$  a ma taki sam priorytet jak b
- $a > b$  a ma większy priorytet niż b

Przykład relacji operatorów

|    | + | - | * | / | id | \$ |
|----|---|---|---|---|----|----|
| +  | > | > | < | < | <  | >  |
| -  | > | > | < | < | <  | >  |
| *  | > | > | > | > | <  | >  |
| /  | > | > | > | > | <  | >  |
| id | > | > | > | > |    | >  |
| \$ | < | < | < | < | <  |    |

Maciek Gębala

Analiza wstępująca, Gramatyki operatorowe, Gramatyki LR(k)

## Analiza dla gramatyki operatorowej

- Weźmy słowo do wyprowadzenia  $\$ id + id * id \$$
- Wstawmy relacje między terminale  $\$ < id > + < id > * < id > \$$
- $\$ < E > + < E > * < E > \$$
- (ignoruj znaki nieterminalne)
- $\$ < + < * > \$$
- $\$ < + > \$$
- $\$ \$$

Maciek Gębala

Analiza wstępująca, Gramatyki operatorowe, Gramatyki LR(k)

## Obliczanie relacji $<, \doteq, >$

- Zbiór  $LEADING(A)$  (pierwsze terminale wyprowadzane z  $A$ ).
  - 1  $a \in LEADING(A)$  jeśli mamy produkcję  $A \rightarrow Ba\beta$  lub  $A \rightarrow a\beta$ .
  - 2 Jeśli istnieje produkcja  $A \rightarrow B\alpha$  i  $a \in LEADING(B)$  to  $a \in LEADING(A)$ .
  - 3 Dla wszystkich nieterminali liczymy 1 i powtarzamy 2 aż nic się nie zmienia.
- Analogicznie  $TRAILING(A)$  (ostatnie terminale wyprowadzane z  $A$ ).

Maciek Gębala

Analiza wstępująca, Gramatyki operatorowe, Gramatyki LR(k)

## Obliczanie relacji $<, \doteq, >$

Dla każdej produkcji  $A \rightarrow x_1 \dots x_k$

- 1 Jeśli  $x_i$  i  $x_{i+1}$  są terminalami to  $x_i \doteq x_{i+1}$ .
- 2 Jeśli  $x_i$  i  $x_{i+2}$  są terminalami a  $x_{i+1}$  nieterminalem to  $x_i \doteq x_{i+2}$ .
- 3 Jeśli  $x_i$  jest terminalem a  $x_{i+1}$  nieterminalem to dla każdego  $a \in LEADING(x_{i+1})$  mamy  $x_i < a$ .
- 4 Jeśli  $x_i$  jest nieterminalem a  $x_{i+1}$  terminalem to dla każdego  $a \in TRAILING(x_i)$  mamy  $a > x_{i+1}$ .

Maciek Gębala

Analiza wstępująca, Gramatyki operatorowe, Gramatyki LR(k)

Notatki

Notatki

Notatki

Notatki

Dla każdego terminala  $a$  definiujemy  $f(a), g(a) \in \mathbb{N}$  tak aby

$$f(a) < g(b) \iff a \prec b$$

$$f(a) = g(b) \iff a \doteq b$$

$$f(a) > g(b) \iff a \succ b$$

Przykład

|     | + | - | * | / | id | \$ |
|-----|---|---|---|---|----|----|
| $f$ | 2 | 2 | 4 | 4 | 6  | 0  |
| $g$ | 1 | 1 | 3 | 3 | 5  | 0  |

## Algorytm analizy metodą pierwszeństwa operatorów

- Wejście: Tekst wejściowy  $w$  i tablica relacji priorytetów.
- Wyjście: Jeśli  $w$  jest poprawny to poprawny ciąg produkcji prawostronnego wyprowadzenia, w p.p. informacje o błędzie.
- Nieterminalne są ignorowane w porównaniach operacjach na stosie a wykorzystywane w redukcjach.

## Algorytm analizy metodą pierwszeństwa operatorów

```
repeat forever
  if  $top(stack) = \$$  and  $first(in) = \$$  then
    return ACCEPT
  else
    if  $top(stack) \prec first(in)$  or  $top(stack) \doteq first(in)$  then
      przenieś  $first(in)$  na  $stack$ 
    else
      if  $!(top(stack) \succ first(in))$  then error()
    else
      while  $top(stack) \succ first(in)$  do /* redukcja */
        pop(stack)
```

## Analiza LR(k)

- Wejście przeglądamy od lewej do prawej ( $L$ ).
- Otrzymujemy wyprowadzenie prawostronne ( $R$ ).
- $k$  - ilość symboli wejścia potrzebnych do podjęcia decyzji. Jeśli  $k = 1$  to jest często pomijane.

Prawie wszystkie języki programowania można opisać gramatyką bezkontekstową typu  $LR(1)$ .

## Algorytm analizy LR

- Analizator LR składa się z wejścia, wyjścia, stosu, programu sterującego i tablicy analizatora.
- Na stosie zapamiętywane są symbole z gramatyki i specjalne symbole zwane **stanami**.
- Symbole stanu podsumowują informacje zawarte na stosie po nim.
- Symbol stanu i aktualny symbol wejściowy są używane do podejmowania decyzji na podstawie tablicy analizatora.

Maciek Gębala

Analiza wstępująca, Gramatyki operatorowe, Gramatyki LR(k)

## Tablica analizatora

- Tablica składa się z dwóch części: akcji i przejść.
- Na podstawie  $s$  - stanu z wierzchołka stosu i  $a$  symbolu na wejściu wykonujemy  $akcja[s, a]$ 
  - ➊ przesunąć, na stos  $a$  i  $s' = akcja[s, a]$ ,
  - ➋ zredukuj zgodnie z produkcją  $A \rightarrow \beta$ , na stos  $A$  i stan wyliczony przez przejście,
  - ➌ akceptuj,
  - ➍ błąd.
- Konfiguracja analizatora LR to zawartość stosu i nieprzetworzone wejście

$(s_0 X_1 s_1 X_2 s_2 \dots X_m s_m, a_i a_{i+1} \dots a_n \$)$

i odpowiada prawostronnej formie wyprowadzenia

$X_1 X_2 \dots X_m a_i a_{i+1} \dots a_n$

Maciek Gębala

Analiza wstępująca, Gramatyki operatorowe, Gramatyki LR(k)

## Działanie analizatora LR

- ➊  $akcja[s_m, a_i] = shift\ s\ (s_0 X_1 s_1 X_2 s_2 \dots X_m s_m, a_i a_{i+1} \dots a_n \$) \vdash (s_0 X_1 s_1 X_2 s_2 \dots X_m s_m a_i s, a_{i+1} \dots a_n \$)$
- ➋  $akcja[s_m, a_i] = A \rightarrow \beta\ (s_0 X_1 s_1 X_2 s_2 \dots X_m s_m, a_i a_{i+1} \dots a_n \$) \vdash (s_0 X_1 s_1 X_2 s_2 \dots X_{m-r} s_{m-r} A s, a_i a_{i+1} \dots a_n \$)$ ,  
gdzie  $s = przejście[s_{m-r}, A]$  i  $\beta = X_{m-r+1} \dots X_m$  ( $r = |\beta|$ ).
- ➌  $akcja[s_m, a_i] = accept$  – koniec analizy.
- ➍  $akcja[s_m, a_i] = error$  – wykryto błąd.

Maciek Gębala

Analiza wstępująca, Gramatyki operatorowe, Gramatyki LR(k)

## Przykład

### Gramatyka

$E \rightarrow E + T\ (1) \mid T\ (2)$   
 $T \rightarrow T * F\ (3) \mid F\ (4)$   
 $F \rightarrow (E)\ (5) \mid ID\ (6)$

### Akcje kodujemy następująco:

- ➊  $s_i$  – przesunięcie  $i$ , wstawiamy na stos stan  $s_i$ ,
- ➋  $r_j$  – redukcja według produkcji o numerze  $j$
- ➌  $acc$  – akceptacja,
- ➍ puste miejsce oznacza błąd.

Maciek Gębala

Analiza wstępująca, Gramatyki operatorowe, Gramatyki LR(k)

### Notatki

### Notatki

### Notatki

### Notatki

### Przykład - Tablica analizatora

| stan     | akcja |       |       |       |          |       | przejście |       |          |
|----------|-------|-------|-------|-------|----------|-------|-----------|-------|----------|
|          | $id$  | $+$   | $*$   | $($   | $)$      | $\$$  | $E$       | $T$   | $F$      |
| $S_0$    | $S_5$ |       |       |       | $S_4$    |       | $S_1$     | $S_2$ | $S_3$    |
| $S_1$    |       | $S_6$ |       |       |          | $acc$ |           |       |          |
| $S_2$    |       | $r_2$ | $S_7$ |       | $r_2$    | $r_2$ |           |       |          |
| $S_3$    |       | $r_4$ | $r_4$ |       | $r_4$    | $r_4$ |           |       |          |
| $S_4$    | $S_5$ |       |       | $S_4$ |          |       | $S_8$     | $S_2$ | $S_3$    |
| $S_5$    |       | $r_6$ | $r_6$ |       | $r_6$    | $r_6$ |           |       |          |
| $S_6$    | $S_5$ |       |       | $S_4$ |          |       |           | $S_9$ | $S_3$    |
| $S_7$    | $S_5$ |       |       | $S_4$ |          |       |           |       | $S_{10}$ |
| $S_8$    |       | $S_6$ |       |       | $S_{11}$ |       |           |       |          |
| $S_9$    |       | $r_1$ | $S_7$ |       | $r_1$    | $r_1$ |           |       |          |
| $S_{10}$ |       | $r_3$ | $r_3$ |       | $r_3$    | $r_3$ |           |       |          |
| $S_{11}$ |       | $r_5$ | $r_5$ |       | $r_5$    | $r_5$ |           |       |          |

Maciek Gębala      Analiza wstępująca. Gramatyki operatorowe. Gramatyki  $LR(k)$

## Przykład

| stos                                   | wejście          | akcja  |
|--|------------------|--|
| $S_0$                                  | $id + id * id^*$ | $S_5$  |
| $S_0 id S_5$                           | $+ id * id^*$    | $r_6 (F \rightarrow id), p[s_0, F] = S_3$    |
| $S_0 F S_3$                            | $+ id * id^*$    | $r_4 (T \rightarrow F), p[s_0, T] = S_2$     |
| $S_0 T S_2$                            | $+ id * id^*$    | $r_2 (E \rightarrow T), p[s_0, E] = S_1$     |
| $S_0 E S_1$                            | $+ id * id^*$    | $S_6$  |
| $S_0 E S_1 + S_6$                      | $id * id^*$      | $S_5$  |
| $S_0 E S_1 + S_6 id S_5$               | $* id^*$         | $r_6 (F \rightarrow id), p[s_6, F] = S_3$    |
| $S_0 E S_1 + S_6 F S_3$                | $* id^*$         | $r_4 (T \rightarrow F), p[s_6, T] = S_9$     |
| $S_0 E S_1 + S_6 T S_9$                | $* id^*$         | $S_7$  |
| $S_0 E S_1 + S_6 T S_9 * S_7$          | $id^*$           | $S_5$  |
| $S_0 E S_1 + S_6 T S_9 * S_7 id S_5$   | $\$$             | $r_6 (F \rightarrow id), p[s_7, F] = S_{10}$ |
| $S_0 E S_1 + S_6 T S_9 * S_7 F S_{10}$ | $\$$             | $r_3 (T \rightarrow T * F), p[s_6, T] = S_9$ |
| $S_0 E S_1 + S_6 T S_9$                | $\$$             | $r_1 (E \rightarrow E + T), p[s_0, E] = S_1$ |
| $S_0 E S_1$                            | $\$$             | $acc$  |

Maciek Gębala Analiza wstępująca. Gramatyki operatorowe. Gramatyki  $LR(k)$ 

## Budowanie tablic analizatorów

- Aby zbudować analizator typu LR wystarczy zbudować tablicę analizatora dla podanej gramatyki.
- Rozpatrzymy trzy metody (typy), od najprostszego i najsłabszego do najskomplikowanego i najsilniejszego:
  - 1 SLR – prosty LR;
  - 2 kanoniczny LR;
  - 3 LALR – podglądające LR.
- Budowa poszczególnych analizatorów na kolejnych wykładach.

Maciek Gębala      Analiza wstępująca. Gramatyki operatorowe. Gramatyki  $LR(k)$

## Notatki

[illegible]

Notatki

[illegible]

## Notatki

This image shows a full page of white paper with ten horizontal dashed lines, typical of primary school handwriting practice paper. The lines are evenly spaced and extend across the entire width of the page. There is no text or other markings on the paper.

Notatki

[illegible]