

Analiza składniowa. Translacja sterowana składnią

Języki formalne i techniki translacji - Wykład 7

Maciek Gębala

22 listopada 2022

Maciek Gębala Analiza składniowa. Translacja sterowana składnią

Wprowadzenie

- Analiza składniowa jest drugą fazą kompilacji (po analizie leksykalnej).
- Analiza składniowa przetwarza dane w postaci **symboli leksykalnych** na **drzewo wyprowadzenia** pewnej gramatyki.
- Sprawdzanie kodu pod względem poprawności składni.
- Gromadzenie dodatkowych informacji, np. w tablicy symboli, potrzebnych dla dalszych kroków kompilacji.

Maciek Gębala Analiza składniowa. Translacja sterowana składnią

Typy analizatorów składniowych

- Zstępujące (**top-down**) – tworzenie drzewa wyprowadzenia od góry.
- Wstępujące (**bottom-up**) – tworzenie drzewa wyprowadzenia od liści w górę do korzenia.
- Metody te działają tylko dla pewnych podklas gramatyk bezkontekstowych (jednoznaczne, deterministyczne).

Maciek Gębala Analiza składniowa. Translacja sterowana składnią

Przykład niejednoznaczności

- Rozważmy gramatykę dla prostych wyrażeń arytmetycznych

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

- Weźmy słowo $id + id * id$
- Mamy dwa różne wyprowadzenia lewostronne

$$E \Rightarrow E + E \Rightarrow id + E \Rightarrow id + E * E$$

$$\Rightarrow id + id * E \Rightarrow id + id * id$$

$$E \Rightarrow E * E \Rightarrow E + E * E \Rightarrow id + E * E$$

$$\Rightarrow id + id * E \Rightarrow id + id * id$$

- Gramatyka jest więc niejednoznaczna.

Maciek Gębala Analiza składniowa. Translacja sterowana składnią

Notatki

Notatki

Notatki

Notatki

Eliminowanie niejednoznaczności

- Niejednoznaczność można czasami wyeliminować.
- Jak zmienić gramatykę

```
instr → if wyr then instr
      | if wyr then instr else instr
      | inna
```

- Gramatyka jednoznaczna

```
instr → p_instr | n_instr
p_instr → if wyr then p_instr else p_instr
        | inna
n_instr → if wyr then instr
        | if wyr then p_instr else instr
```

- Porównaj wyprowadzenia dla $if E_1 then if E_2 then I_1 else I_2$

Maciek Gębala

Analiza składniowa, Translacja sterowana składnią

Eliminacja lewostronnej rekurencji

Gramatyka jest **lewostronnie rekurencyjna** jeśli istnieje nieterminal A taki, że istnieje wyprowadzenie $A \Rightarrow^+ A\alpha$ dla pewnego α .
Metody zstępujące (**up-down**) nie dają się zastosować do takich gramatyk.

Lewostronną rekurencję można wyeliminować.

Maciek Gębala

Analiza składniowa, Translacja sterowana składnią

Obsługa błędów

W przypadku wystąpienia błędu wypisujemy go (czytelnie z podaniem miejsca wystąpienia) i staramy się kontynuować przetwarzanie.

Strategie odzyskiwania kontroli

- tryb paniki
- poziom frazy
- produkcje dla błędów
- korekta globalna

Maciek Gębala

Analiza składniowa, Translacja sterowana składnią

Tryb paniki

- Sposób najprostszy w implementacji.
- Po natrafieniu na błąd usuwamy kolejne symbole aż trafimy na taki który umożliwi nam ponowną synchronizację danych z analizatorem (np. ograniczniki instrukcji).
- Mimo ominięcia pewnej liczby symboli metoda pozwala łatwo kontynuować pracę.

Maciek Gębala

Analiza składniowa, Translacja sterowana składnią

Notatki

Notatki

Notatki

Notatki

Poziom frazy

- Maciek Gębala Analiza składniowa. Translacja sterowana składnią

Produkcje dla błędów

- Maciek Gębala Analiza składniowa. Translacja sterowana składnią

Korekta globalna

- Maciek Gębala Analiza składniowa. Translacja sterowana składnią

Translacja sterowana składnią

Maciek Gębala Analiza składniowa. Translacja sterowana składnią

[illegible][illegible]

This image shows a single sheet of white paper with horizontal blue ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins or other markings on the paper.

[illegible]

Definicje

- **Definicje sterowane składnią** – ukrywają wiele szczegółów implementacyjnych, nie wymagają jawnego określania kolejności obliczania reguł semantycznych.
- **Schematy translacji** – wskazują kolejność wyliczania reguł semantycznych, pokazują więcej szczegółów implementacyjnych.

Maciek Gębala

Analiza składniowa. Translacja sterowana składnią

Definicje sterowane składnią

- Definicje sterowane składnią są uogólnieniem gramatyki bezkontekstowej przez związanie z każdym symbolem pewnego zbioru atrybutów.
- Atrybuty dzielimy na syntetyzowane i dziedziczone.
- Wartości atrybutów w węźle drzewa wyprowadzenia są określane przez reguły semantyczne związane z produkcją przypisaną do tego węzła.

Maciek Gębala

Analiza składniowa. Translacja sterowana składnią

Atrybuty

Z każdą produkcją gramatyki $B \rightarrow X_1 X_2 \dots X_n$ wiążemy zbiór reguł semantycznych postaci $b \leftarrow f(p_1, p_2, \dots, p_k)$.

- b jest atrybutem syntetyzowanym symbolu B a p_1, p_2, \dots, p_k są atrybutami symboli X_1, X_2, \dots, X_n .
- b jest atrybutem dziedziczonym symbolu X_i a p_1, p_2, \dots, p_k są atrybutami symboli B, X_1, X_2, \dots, X_n .

Maciek Gębala

Analiza składniowa. Translacja sterowana składnią

Definicje S-atrybutowe i L-atrybutowe

Definicje S-atrybutowe

Gramatyka posiada tylko atrybuty syntetyzowane.

Definicje L-atrybutowe

Gramatyka posiada tylko atrybuty syntetyzowane lub atrybuty dziedziczone symbolu X_i w produkcji $B \rightarrow X_1 X_2 \dots X_n$ które zależą tylko od atrybutów symboli X_1, X_2, \dots, X_{i-1} oraz atrybutu dziedziczonych symbolu B .

Każda definicja S-atrybutowa jest definicją L-atrybutową.

Maciek Gębala

Analiza składniowa. Translacja sterowana składnią

Notatki

Notatki

Notatki

Notatki

Produkcja	Reguły semantyczne
$D \rightarrow TL$	$L.dz \leftarrow T.typ$
$T \rightarrow int$	$T.typ \leftarrow integer$
$T \rightarrow real$	$T.typ \leftarrow real$
$L \rightarrow L_1, id$	$L_1.dz \leftarrow L.dz; id.typ \leftarrow L.dz$
$L \rightarrow id$	$id.typ \leftarrow L.dz$

Generator analizatorów wstępujących - Bison (Yacc)

- Bison dla zadanej specyfikacji generuje kod źródłowy analizatora składniowego.
- Program Bison w łatwy sposób może współpracować z generatorem analizatorów leksykalnych Flex.
- Program generuje analizator redukujący LALR.
- Nieterminal lewej strony pierwszej produkcji jest domyślnie symbolem startowym.
- W razie konfliktu redukcja/redukcja wybierana jest akcja wynikająca z kolejności zapisu.
- W razie konfliktu redukcja/przesunięcie wybierane jest przesunięcie.

Bison - plik specyfikacji

- Funkcja parsująca `yyparse()`.
- Funkcja zwracająca błędy `yyerror()`.
- Plik specyfikacji składa się z trzech sekcji (rozdzielonych `%`):
 - 1 sekcja definicji,
 - 2 sekcja reguł przetwarzania,
 - 3 sekcja podprogramów.
- W sekcji definicji definiujemy tokeny (terminale) używane w gramatyce.
- Reguła przetwarzania składa się z produkcji i operacji (w języku C).

Współpraca z LEX-em

- Parser korzysta z funkcji `int yylex()` do czytania tokenów którą można napisać samemu lub skorzystać z LEX-a.
- Tokeny użyte w parserze mogą być wyeksportowane do pliku nagłówkowego.
- Również wartość tokena może być przeniesiona do parsera – zmienna globalna `yyval`. Domyślny typ to `int` ale można zmienić na kilka typów posługując się unią i deklaracjami typów.

Przykład

Gramatyka

$$\begin{aligned} E &\rightarrow E \text{ or } T \mid T \\ T &\rightarrow T \text{ and } F \mid F \\ F &\rightarrow \text{not } F \mid (E) \mid \text{true} \mid \text{false} \end{aligned}$$

Pliki

- `bool-calc.y`
- `bool-calc.l`

Notatki

Translacja sterowana składnią w BISON-ie

- Reguły semantyczne w BISON-ie są obliczane w trakcie translacji.
- Drzewo wyprowadzenia nie jest jawnie konstruowane. Porządek wyliczania atrybutów jest narzucony przez samą metodę analizy.
- **Zalety:** prostota i efektywność translatora.
- **Wady:** możliwość przetwarzania tylko L-atrybutowych definicji sterowanych składnią.
- Realizując translację sterowaną składnią w BISON-ie bezpieczniej jest korzystać z mechanizmu atrybutów wbudowanego w generator.
- Używanie zmiennych globalnych może mieć efekty uboczne i nie gwarantuje prawidłowej kolejności akcji wykonywanych przez LR-parser.

Notatki

Atrybuty – ogólne reguły

- Z każdym symbolem w produkcji związany jest atrybut

$$\begin{array}{ccccccc} A & : & X_1 & X_2 & X_3 & \dots & X_n \\ \downarrow & & \downarrow & \downarrow & \downarrow & & \downarrow \\ \$\$ & & \$1 & \$2 & \$3 & & \$n \end{array}$$

- Wartość atrybutu jednostki leksykalnej jest nadawana w analizatorze leksykalnym za pomocą zmiennej `yylval`.
- Zmienna `yylval` jest domyślnie typu całkowitego (`int`).
- Akcje semantyczne możemy zapisać z wykorzystaniem atrybutów.

```
E : E '+' num { $$=$1+$3; }
    | num { $$=$1; };
```

Notatki

Atrybuty – ogólne reguły

- Domyślny typ atrybutów można zmodyfikować za pomocą słowa kluczowego `%union` umieszczanego w specyfikacji parsera.
- Wewnątrz deklaracji `%union` można umieścić wszystkie potrzebne typy atrybutów.
- W celu zadeklarowania typów symboli należy najpierw zdefiniować typ atrybutów za pomocą `%union np.`
`%union{ char *text; int ival; }`
a następnie na podstawie tej definicji przypisać typ dla symboli terminalnych
`%token <text> ident`
i dla nieterminalnych
`%type <ival> A`

Notatki

- Zmiana typów w parserze wymusza zmianę typów w skanerze:
dodanie definicji interfejsu
`#include "y.tab.h"`
i odpowiednie nadawanie wartości tokenom
`yyval.ival=atoi(ytext);`

Przykład – Gramatyka S-atrybutowa

```
%union{ char *text; }
%type <text> T L
%token <text> id
%token <text> decl_real decl_int
%%
D : L ',' { YYACCEPT; }
;
L : T id {printf("%s %s;\n",$1,$2);free($2);$=$1;}
  | L ',' id
    {printf("%s %s;\n",$1,$3);free($3);$=$1;}
;
T : decl_int { $$ = "int"; }
  | decl_char { $$ = "char"; }
;
%%
```

Przykład – lekser

```
%option noyywrap
%{
#include "y.tab.h"
%}
id [_a-zA-Z][_a-zA-Z0-9]*
%%
"int" { return decl_int; }
"char" { return decl_char; }
{id} { yyval.text=strdup(yytext); return id; }
";" { return ' '; }
", " { return ' '; }
\n
.
%%
```

Atrybuty dziedziczone

- W BISON-ie można korzystać z atrybutów dziedniczonych, ale definicje muszą być L-trybutowe.
- Wszystkie akcje związane z obliczeniem atrybutów nieterminali z których korzystamy muszą być już wykonane i znajdować się na stosie.
- Atrybuty dziedziczone w BISON-ie oznaczane są niedodatnimi indeksami, w kolejności od szczytu stosu (0) w głąb.
- Typ pola unii do którego się odwołujemy powinien być wskazany (w nawiasach trójkątnych).
- Nie jest przeprowadzana kontrola poprawności sięgania po atrybuty (na stosie).

This image shows a full page of white paper with ten horizontal dashed lines, evenly spaced from top to bottom. There are no margins, text, or other markings on the page.

Notatki

[illegible]

Notatki

[illegible]

Notatki

[illegible]