Документация по проекту «Компилятор языка Паскаль» Подготовили Елфимова Екатерина, Мещеряков Илья и Стребкова Ольга

Синтаксический анализ

Компилятор языка Паскаль, написанный на языке python с испоьзованием библиотеки pyparsing.

На даммный момент в проекте реализованно построение ast-дерева для кода на языке Паскаль.

Структура проекта:

Класс PascalGrammar который отвечает за описание грамматики языка Паскаль. В нем присутствует метод _make_parser грамматику языка и правила ее разбора. Здесь мы возспользовались модулем pyparsing для создания грамматики, используя предоставляемую модулем библиотеку классов для построения грамматики непосредственно в коде Python.

main.py используется для вызова функции.

В файле nodes.py описаны узлы для постороения ast дерева:

AstNode - класс предок для всех последующих классов, определяет для узлов такие свойства как получние потомков узла, и рисование узла

LiteralNode - узел для описания литерала

IdentNode - узел для описания идентификатора

ArrayIdentNode - узел для описания элементов массива

BinOpNode - узел для описания бинарной операции

StmtNode - узел для описания выражения

IdentListNode - узел для описания списка идентификаторов

TypeSpecNode - узел для описания типа

VarDeclNode - узел для описания объявления переменной

ArrayDeclNode - узел для описания объявления массива

VarsDeclNode - узел для описания объявлений переменных

CallNode - узел для описания вызова функции или процедуры

AssignNode - узел для описания присваивания переменной значения

IfNode - узел для описания условного оператора if

WhileNode - узел для описания цикла while

ForNode - узел для описания цикла for

StmtListNode - узел для описания списка выражений

BodyNode - узел для описания тела (внутренности между begin и end) содержащий список выражений

ParamsNode - узел для описания параметров функции/процедуры

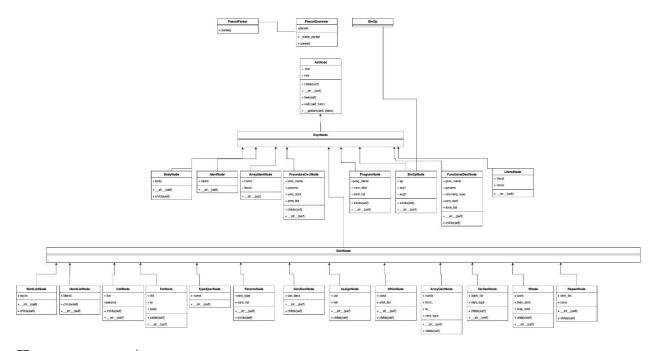
ProgramNode - узел для описания программы

ProcedureDeclNode - узел для описания процедуры

FunctionDeclNode- узел для описания функции

Так же в коде проекта присутствуют комментарии с пояснениями по ходу выполнения программы.

Диаграмма классов, описывающая структуру проекта:



Примеры вход/выход:

вход 1: Program prog1; var

```
j: integer;
BEGIN
    j:=9;
END.
выход 1:
Program
| prog1
- var
| L var_dec
| | idents
| | <sup>L</sup> j
| L integer
L Body
  L ...
    L :=
      ŀj
     L 9 (int)
вход 2:
Program prog1;
var
    k, d: integer;
    j: char;
    g, c: array [1 .. 100] of integer;
BEGIN
    while (i>3) do
       a:=k+2;
    a:=j mod k;
    while (i>3) do
    begin
        a:=k+2;
        a:=j mod k;
    end;
END.
выход 2:
Program
| prog1
- var
| | var_dec
| \ | \ | idents
| | | | k
| | L integer
```

```
| | var_dec
| \ | \ | idents
| | L char
| L arr_decl
| | integer
 | idents
| | <del>|</del> | g
 | <sup>L</sup> c
|  | 1 (int)
| L 100 (int)
L Body
 L ...
   - while
   | | >
   | L :=
   ⊦ k
       L 2 (int)
   -:=
   | L mod
   |  | j
   ^{\mathsf{L}} while
     | >
     | L 3 (int)
     L ...
       | :=
      | L +
       | L 2 (int)
      L :=
        ├ a
        L \mod
         ŀj
         L k
вход 3:
```

```
Program prog1;
var
k, d: integer;
```

```
function t(j:integer; k: char):integer;
   var
       d: integer;
   begin
       for (i:=2 to 0 ) do
          g:=0;
       s:=0;
       if (k>2) then
       begin
          f:=9;
           h:=f;
       end;
       else
          v:=3;
   end;
BEGIN
   t(1,3);
END.
выход 3:
Program
| prog1
- var
| | var_dec
| | | idents
| | L integer
| L function
   | t
   | params
   | | integer
   | | char
   | L j
   - integer
   - var
   | L var_dec
   | | idents
   | | L d
   | L integer
   <sup>L</sup> Body
     L ...
       - for
       | | :=
```

```
| L :=
         ⊦g
        | L 0 (int)
        |- :=
        | ^{L} 0 (int)
        L if
         | L 2 (int)
          | L :=
           ⊦ h
         | L f
         L :=
           ├ v
           L 3 (int)
L Body
  L ...
    ^{\mathsf{L}} call
     | t
     ├ 1 (int)
      L<sub>3</sub> (int)
вход 4:
Program prog1;
var
   procedure t(j:integer; k: char);
       var
           d: integer;
           procedure j(d: integer);
           var
               h:integer;
           begin
               d:=h;
           end;
       begin
           a:=78;
       end;
BEGIN
   if (k>2 \text{ or } 1<0) then
```

```
begin
        f:=9;
        if (k>2 and h>=0) then
            f:=4;
    end;
END.
выход 4:
Program
| prog1
├ var
| L procedure
    | t
    | params
    | | integer
    | | char
    | <sup>L</sup> j
    - var
    | - var_dec
    | | L integer
    | L procedure
         ŀj
         | params
        | | integer
        | L d
         - var
        | L var_dec
        | | idents
         | L integer
         <sup>L</sup> Body
           L ...
            L :=
               ├ d
               ^{\mathsf{L}} h
    <sup>L</sup> Body
      L ...
        L :=
          L 78 (int)
<sup>L</sup> Body
  L ...
    ^{\mathsf{L}} if
      - or
```

```
| | >
     \mid \; \mid \; ^{L} 2 (int)
     | L <
     | | 1
     | L 0 (int)
     L ...
       | :=
       | L 9 (int)
       L if
         - and
         | | >
         | L >=
         | L 0 (int)
         L :=
           | f
           L 4 (int)
вход 5:
Program t;
   k, d: integer;
   j: char;
   g, c: array [1 .. 100] of integer;
function t(j:integer; k: char):integer;
   d: integer;
a:= 0;
procedure t;
   d: integer;
a:=78;
   g: integer;
BEGIN
g[0]:=10;
g[1]:=c[0];
writeln(a, 3, "df", 7+9);
for ( i:=2 to 10 ) do
```

var

var

begin

end;

var

begin

end;

begin

```
k:=2 mod 3;
  l:=j div 4;
  l:=i+8;
  k:=0;
end;
while (i>3) do
  a:=k+2;
for (i:=2 to 6 ) do
   g:=0;
   s:=0;
if (k>2 \text{ and } j>=2) then
  begin
  f:=9;
   end;
else
  v:=3;
END.
выход 5:
Program
| t
- var
| | var_dec
| \ | \ | idents
| | var_dec
| | | idents
| | L char
| | arr_decl
| | | integer
| | | idents
| | function
| | <del>|</del> t
| | | integer
```

```
| | | L var_dec
| | | Linteger
|  | L Body
| | L ...
| | procedure
| | <del>|</del> t
| | | ...
| | | L var_dec
|\ |\ |\ |\ L\ d
| | | Linteger
|  | L Body
| | L :=
| | | | a
| L var_dec
| | idents
| | L g
| Linteger
L Body
 L ...
  | :=
  | <sup>L</sup> 10 (int)
  | :=
  | L c [0 (int)]
  - call
  | | writeln
  | | "df" (str)
  | L +
  | L 9 (int)
  - for
  | | :=
  | L ...
```

```
| :=
 | L mod
 | L 3 (int)
  - :=
 | | 1
  | L div
  |  | j
 | L 4 (int)
  |- :=
  | | 1
  | L +
  | L 8 (int)
 L :=
- while
\mid \; \mid \; ^{L} 3 (int)
| L :=
L +
- for
| | :=
| L :=
| L 0 (int)
-:=
| L 0 (int)
L if
 and
 | |->
 | L >=
 | L 2 (int)
 | ...
 | L :=
```

Документация по проекту «Компилятор языка Паскаль». Подготовили Елфимова Екатерина, Мещеряков Илья и Стребкова Ольга

Семантический анализ

Для проведения семантического анализа были в программу были добавлены файлы semantic.py и symbols.py для определения логики необходимой при проведении семантического анализа.

Рассмотрим файл symbols.py. В данном файле определены классы Symbol и его наследники - BuiltinTypeSymbol, BuiltinFunction, ArraySymbol, VarSymbol, BlockSymbol, ProcedureSymbol, FunctionSymbol. Данные классы описывают структуру соответствующих символов, которые будут записываться в символьную таблицу, необходимую для проверки семантики программы. Классы BuiltinTypeSymbol и BuiltinFunction описывают встроенные типы и функции соответственно.

Рассмотрим файл semantic.py.

Файл содержит класс ScopedSymbolTable, который описывает символьную таблицу. Данный класс содержит набор символов данной таблицы, имя таблицы, уровень вложенности данной таблицы и ссылку на вышестоящую таблицу. В классе определены методы для записи в таблицу нового символа (define) и для поиска в таблице символа (lookup).

Файл содержит класс NodeVisitor, необходимый для совершения обхода дерева и проверки семантики. Его потомком является класс SemanticAnalyzer, в котором присутствует описание методов для посещения конкретных типов узлов:

- visit_BinOpNode
- visit_IdentNode
- visit_LiteralNode
- visit_ProgramNode
- visit_VarsDeclNode

- visit_VarDeclNode
- visit_ArrayDeclNode
- visit_ArrayIdentNode
- visit_BodyNode
- visit_StmtListNode
- visit_AssignNode
- visit_ProcedureDeclNode
- visit_FunctionDeclNode
- visit_CallNode
- visit_IfNode
- visit_WhileNode
- visit_ForNode

Для проверки семантики программы в классе присутствует поле current_scope, в котором хранится текущий уровень символьной таблицы. Благодаря такому построению структуры у нас есть возможность создавать несколько вложенных областей (в языке Pascal это вложенные процедуры и функции).

Кроме того, класс реализует логику для проверки типов для бинарных операций. Для этого в классе имеется поле BinOpArgTypes, в котором описано соответствие бинарных операций и допустимых для них типов.

Таким образом, в рамках семантического анализа нами была реализована логика по проверки совместимости типов, а также проверки объявлений переменных, функций и процедур.

Пример построения символьной таблицы.

```
Program t;
var
g : integer;
c : boolean;
a : char;
d,n: array [1 .. 100] of integer;
h : array [1 .. 7] of boolean;
```

```
function Alpha(a,b: integer, g:char):integer;
var y: integer;
begin
g:='c';
c:=True;
ReadLn(g);
if (c) then
c:=False;
for (a:=2 to 6) do
begin
end;
end;
BEGIN
h[1]:=d[2]=1;
g:=g;
END.
Program
- t
 - var
  - var dec
     - idents
      Lg
     L integer
    var dec
     - idents
       L<sub>C</sub>
     L boolean
    var dec
     - idents
      L<sub>a</sub>
     L char
    arr decl
     - integer
      - idents
       \begin{bmatrix} d \\ n \end{bmatrix}
     - 1 (int)
     L 100 (int)
    arr decl
     - boolean
       idents
      L h
     1 (int)
7 (int)
   L function
     - Alpha
       params
       - var dec
          - idents
           l a
b
         Linteger
       L var_dec
          - idents
           Lg
          L char
       integer
       var
       L var_dec
         - idents
           Ly
          Linteger
```

SCOPE (SCOPED SYMBOL TABLE)

Scope name : Alpha
Scope level : 2
Enclosing scope: global

```
SCOPE (SCOPED SYMBOL TABLE)
_____
Scope name : global
Scope level : 1
Enclosing scope: None
Scope (Scoped symbol table) contents
______
integer: 'integer'
   char: 'char'
boolean: 'boolean'
   Read: 'Read'
ReadLn: 'ReadLn'
 Write: 'Write'
WriteLn: 'WriteLn'
      g: '<g:integer>'
      c: '<c:boolean>'
      a: '<a:char>'
      d: '<d:integer[1 .. 100]>'
      n: '<n:integer[1 .. 100]>'
      h: '<h:boolean[1 .. 7]>'
  Alpha: "<FunctionSymbol(name=Alpha, parameters=['a: integer', 'b: integer',
'g: char']): integer>"
```

Scope (Scoped symbol table) contents

integer: 'integer'
 char: 'char'
boolean: 'boolean'
 Read: 'Read'
ReadLn: 'ReadLn'
 Write: 'Write'
WriteLn: 'WriteLn'

a: '<a:integer>'
b: '<b:integer>'
g: '<g:char>'

Кодогенерация (B java byte code)

Выполнена Елфимовой Екатериной

Для реализации кодогенерации в программе добавлены инструкции, собирающие код по исходной поданной программе на языке Pascal. Программа генерирует ассемблер для виртуальной машины Java (Jasmin). После генерации кода необходимо с помощью сборщика Jasmin скомпилировать ассемблер сгенерированный программой в java byte code. Для этого необходимо выполнить:

java -jar jasmin.jar target_file.j

После выполнения данной команды будет target_file.j будет скомпилирован в файл .class, который может быть исполнен JVM.

Для исполнения полученного файла необходимо выполнить команду: java target_file

Для кодогенерации ассемблера были изучена структура такой программы и выделены некоторые инструкции, которые были использованы при генерации кода:

- $ldc\ 0$ кладет на стек значение 0
- iload_1 берет из локальной переменной с адресом 1 значение и кладет его на стек (префикс i – целое число)
- *istore_1* снимает со стека значение и сохраняет его в локальную переменную с адресом 1 (префикс і целое число)

Бинарные операции, которые снимают со стека два значения производят соответствующую операцию и возвращают результат на стек:

- iadd
- isub
- imul
- idiv
- iand

• ior

Префикс і обозначает что операция будет выполнена над целыми числами.

Ключевое слово Program из языка Pascal интерпретируется как класс. Для этого прописывается:

.class public ClassName

.super java/lang/Object

для описания класса и вызова конструктора суперкласса.

Для описания методов используются конструкции вида:

.method public static MethName(II)V

.limit stack 10

.limit locals 10

//some code

return

.end method

Здесь приведено описание процедуры. Символы II говорят о том, что метод будет принимать два параметра целочисленного типа. Символ V описывает тип возвращаемого значения, в данном случае — void. Limit stack и limit locals заявляют выделяемый размер стека и максимальное количество локальных переменных.

Для вызова методов используется инструкция:

invokestatic ClassName/MethName(II)V

Для вызова библиотечных функций:

- getstatic java/lang/System/in Ljava/io/InputStream;
 invokevirtual java/io/InputStream/read()I (для ввода)
- invokevirtual java/io/PrintStream/println(I)V
 getstatic java/lang/System/out Ljava/io/PrintStream; (для вывода)

Глобальные переменные программы Pascal интерпретируются как

статические поля класса. Их объявление выглядит так:

.field public static field1 I

символ I указывает тип поля, в данном случае поле целочисленное.

Для операций с данным полем используются инструкции:

- putstatic ClassName/field1 I снимает значение со стека и записывает его в поле
- getstatic ClassName/field1 I кладет значение из поля в стек

Для описания условного оператора и цикла были использованы такие конструкции как label и инструкция goto. Label представляет из себя инструкцию к которой можно вернуться с помощью инструкции goto some_label. Кроме того использовались инструкции для проверки условия:

- if_icmpeq
- if_icmpne
- if_icmge
- if_icmle
- if_icmgt
- if_icmlt

Массивы описываются с помощью следующей инструкции:

.field public static mass [I

После чего они должны быть проинициализированы, что исполняется с помощью инструкций:

ldc 100 //100-число эл-ов массива

newarray int //инициализация

putstatic ClassName/mass [I //coxpанение в соответствующее поле

Тело программы Pascal исполняется в методе *main*, который является точкой входа в программу.

Примеры генерации:

Вход 1:

```
Program pr1;
            var
            g, b : integer;
            d, n: array [1 .. 100] of integer;
            function Alpha(a: integer):integer;
            var y: integer;
            begin
            a:=a*10;
            end;
            BEGIN
            g := 0;
            while (g<10) do
            begin
                d[1] := g;
                Write(d[1]);
                 g := g+1;
            end;
            g:=Alpha(4);
            WriteLn(g);
            END.
```

Выход 1:

```
.class public pr1
.super java/lang/Object
.field public static g I
.field public static b I
.field public static d [I
.field public static n [I
.method public static Alpha(I)I
.limit stack 100
.limit locals 100
iload 0
1dc 10
imul
istore 0
iload 0
ireturn
.end method
.method
                         public static main([Ljava/lang/String;)V
                     100
.limit stack
.limit locals
                      100
ldc 100
newarray int
putstatic pr1/d [I
ldc 100
newarray int
putstatic pr1/n [I
ldc 0
putstatic pr1/g I
while 0:
```

```
getstatic pr1/g I
1dc 10
swap
if icmplt done 0
getstatic pr1/d [I
ldc 1
getstatic pr1/g I
iastore
getstatic
                      java/lang/System/out Ljava/io/PrintStream;
getstatic pr1/d [I
ldc 1
iaload
                      java/io/PrintStream/println(I) V
invokevirtual
getstatic pr1/g I
ldc 1
iadd
putstatic pr1/g I
goto while 0
done 0:
ldc 4
invokestatic pr1/Alpha(I)I
putstatic pr1/g I
                      java/lang/System/out Ljava/io/PrintStream;
getstatic
getstatic pr1/g I
                      java/io/PrintStream/println(I)V
invokevirtual
  return
.end method
      Вход 2:
```

```
Program pr2;
        var
        g, b : integer;
        d,n: array [1 .. 100] of integer;
        BEGIN
        g:=4*4/2+2-5;
        b := 0;
        WriteLn(g);
        WriteLn(b);
        if (b>g) then
             b := 0;
        else
            b := 1;
        WriteLn(b);
        while (b<q) do
             b := b+1;
             WriteLn(b);
        end;
        END.
```

Выход 2:

```
.class public pr2
.super java/lang/Object
.field public static g I
.field public static b I
.field public static d [I
.field public static n [I
.method public static main([Ljava/lang/String;)V
```

```
.limit stack
                      100
.limit locals
                       100
ldc 100
newarray int
putstatic pr2/d [I
ldc 100
newarray int
putstatic pr2/n [I
ldc 4
ldc 4
imul
ldc 2
idiv
ldc 2
iadd
ldc 5
isub
putstatic pr2/g I
ldc 0
putstatic pr2/b I
                       java/lang/System/out Ljava/io/PrintStream;
getstatic
getstatic pr2/g I
                       java/io/PrintStream/println(I)V
invokevirtual
                       java/lang/System/out Ljava/io/PrintStream;
getstatic
getstatic pr2/b I
                       java/io/PrintStream/println(I) V
invokevirtual
getstatic pr2/b I
getstatic pr2/g I
swap
if icmpgt else 0
ldc 0
putstatic pr2/b I
\verb"goto endif_0"
else 0:
ldc 1
putstatic pr2/b I
endif 0:
                       java/lang/System/out Ljava/io/PrintStream;
getstatic
getstatic pr2/b I
                       java/io/PrintStream/println(I) V
invokevirtual
while 1:
getstatic pr2/b I
getstatic pr2/g I
swap
if icmplt done 1
getstatic pr2/b I
ldc 1
iadd
putstatic pr2/b I
                       java/lang/System/out Ljava/io/PrintStream;
getstatic
getstatic pr2/b I
invokevirtual
                       java/io/PrintStream/println(I) V
goto while 1
done 1:
   return
.end method
```

Вход 3:

```
Program pr3;
var x, y, i: integer;
function Alpha(a,b: integer):integer;
```

```
var y: integer;
begin
a:=b+a;
end;
BEGIN
x:=1;
y := x;
y := (y+45) * (1+1);
x := 3 + y - 1;
Write(y);
y:=Alpha(3,5);
while (y<100) do
begin
    Write(y);
    y := y+1;
    Write(y);
if (y>1) then
    x := 1;
else
    x := 2;
END.
      Выход 3:
.class public pr3
.super java/lang/Object
.field public static x I
.field public static y I
.field public static i {\tt I}
.method public static Alpha(II)I
.limit stack 100
.limit locals 100
iload_1
iload 0
iadd
istore 0
iload 0
ireturn
.end method
                           public static main([Ljava/lang/String;)V
.method
                      100
.limit stack
                      100
.limit locals
ldc 1
putstatic pr3/x I
getstatic pr3/x I
putstatic pr3/y I
getstatic pr3/y I
1dc 45
iadd
ldc 1
ldc 1
iadd
imul
```

java/lang/System/out Ljava/io/PrintStream;

putstatic pr3/y I

getstatic pr3/y I

putstatic pr3/x I

ldc 3

iadd
ldc 1
isub

getstatic

```
getstatic pr3/y I
invokevirtual
                       java/io/PrintStream/println(I)V
ldc 3
ldc 5
invokestatic pr3/Alpha(II)I
putstatic pr3/y I
while_0:
getstatic pr3/y I
ldc 100
swap
if icmplt done 0
                       java/lang/System/out Ljava/io/PrintStream;
getstatic
getstatic pr3/y I
                       java/io/PrintStream/println(I) V
invokevirtual
getstatic pr3/y I
ldc 1
iadd
putstatic pr3/y I
                       java/lang/System/out Ljava/io/PrintStream;
getstatic
getstatic pr3/y I
invokevirtual
                       java/io/PrintStream/println(I)V
goto while 0
done 0:
getstatic pr3/y I
ldc 1
swap
if_icmpgt else_1
ldc 1
putstatic pr3/x I
goto endif 1
else 1:
1dc \overline{2}
putstatic pr3/x I
endif 1:
  return
.end method
```