|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Министерство науки и высшего образования РФ | | | | | | | | |
|  | | |  | | |  | | |
| Фгбоу впо «пермский государственный национальный исследовательский университет» | | | | | | | | |
|  | | |  | | |  | | |
|  | | Отчет  по реализации компилятора для подмножества языка Паскаль | | | | |  | |
|  | | |  | | |  | | |
|  | Работу выполнили  студент гр. ПМИ-1,2  Захаров И.В. \_\_\_\_\_\_\_\_\_  «\_\_\_» \_\_\_\_\_\_\_\_\_\_ 2021 | | |  | Проверил  ассистент кафедры МОВС  Пономарев Ф.А. \_\_\_\_\_\_\_\_\_  «\_\_\_» \_\_\_\_\_\_\_\_\_\_ 2021 | | |  |
|  | | |  | | |  | | |
|  | | | Пермь 2021 | | |  | | |

содержание

[1 Постановка задачи 5](#_Toc91103434)

[2 Модуль ввода-вывода 7](#_Toc91103435)

[2.1 Описание 7](#_Toc91103436)

[2.2 Проектирование 7](#_Toc91103437)

[2.3 Реализация 8](#_Toc91103438)

[2.4 Тестирование 11](#_Toc91103439)

[2.4.1 Создание из строки исходного кода 11](#_Toc91103440)

[2.4.2 Создание из файла 11](#_Toc91103441)

[2.4.3 Получение текущего символа, позиции 12](#_Toc91103442)

[2.4.4 Вывод сообщения об ошибке 12](#_Toc91103443)

[3 Лексический анализатор 13](#_Toc91103444)

[3.1 Описание 13](#_Toc91103445)

[3.2 Проектирование 13](#_Toc91103446)

[3.3 Реализация 14](#_Toc91103447)

[3.3.1 Лексический анализатор 14](#_Toc91103448)

[3.3.2 Токен 18](#_Toc91103449)

[3.3.3 Токен-константа 18](#_Toc91103450)

[3.3.4 Токен-идентификатор 18](#_Toc91103451)

[3.3.5 Токен-оператор 19](#_Toc91103452)

[3.3.6 Хэш-таблицы и перечислимый тип для операторов 19](#_Toc91103453)

[3.4 Тестирование 20](#_Toc91103454)

[4 Синтаксический анализатор 22](#_Toc91103455)

[4.1 Описание 22](#_Toc91103456)

[4.2 Проектирование 22](#_Toc91103457)

[4.3 Реализация 23](#_Toc91103458)

[4.3.1 Функция next\_token 23](#_Toc91103459)

[4.3.2 Функция accept 23](#_Toc91103460)

[4.3.3 Функции реализующие БНФ 25](#_Toc91103461)

[4.3.4 Полное описание класса 30](#_Toc91103462)

[4.4 Тестирование 38](#_Toc91103463)

[5 Семантический анализатор 40](#_Toc91103464)

[5.1 Описание 40](#_Toc91103465)

[5.2 Проектирование 40](#_Toc91103466)

[5.3 Реализация 40](#_Toc91103467)

[5.3.1 Функция добавления переменной 40](#_Toc91103468)

[5.3.2 Функция приведения типов 41](#_Toc91103469)

[5.3.3 Полное описание класса модуля семантического анализатора 42](#_Toc91103470)

[5.3.4 Описание класса типов 52](#_Toc91103471)

[5.4 Тестирование 54](#_Toc91103472)

[6 Транслятор 57](#_Toc91103473)

[6.1 Описание 57](#_Toc91103474)

[6.2 Проектирование 57](#_Toc91103475)

[6.3 Реализация 58](#_Toc91103476)

[6.3.1 Функция добавления переменной 58](#_Toc91103477)

[6.3.2 Функция для вызова инструкции исходя из переданного оператора 59](#_Toc91103478)

[6.3.3 Функция вызова метода вывода в консоль(writeln) 60](#_Toc91103479)

[6.3.4 Функция вызова метода ввода из консоли(readln) 60](#_Toc91103480)

[6.3.5 Функция для смены знака значения на стеке 61](#_Toc91103481)

[6.3.6 Функция для загрузки константного значения в стек 61](#_Toc91103482)

[6.3.7 Полный код модуля транслятора 61](#_Toc91103483)

[6.4 Тестирование 76](#_Toc91103484)

[6.4.1 Ввод и вывод числа 76](#_Toc91103485)

[6.4.2 Проверка введенного числа на простоту 77](#_Toc91103486)

[6.4.2 Нахождение корня для полинома третьей степени 79](#_Toc91103487)

1. Постановка задачи

Глобальное задание: написать компилятор для подмножества языка Паскаль.

Задание разбивается на отдельные этапы:

1. Модуль ввода-вывода (8 баллов, оценивается совместно с лексическим анализатором).

2. Лексический анализатор (12 баллов, оценивается совместно с модулем ввода-вывода).

3. Синтаксический анализатор (12 баллов) с нейтрализацией синтаксических ошибок (8

баллов).

4. Семантический анализатор с нейтрализацией семантических ошибок (20 баллов).

5. Генерация кода (25 баллов).

Реализовать следующие разделы программы:

Основные разделы программы: раздел описания переменных, раздел операторов. Переменные стандартных типов (Boolean, integer, real, char). Числовые константы. Арифметическое выражение (в выражении допустимы только константы, переменные, операции +, –, \*, / и скобки). Оператор присваивания и составной оператор.

Раздел описания типов. Выражение (полностью, включая арифметические, логические операции, сравнения и т.д., но только над константами и простыми переменными (не индексированные, не поля записи, не указатели)). Условный оператор (if). Оператор цикла с предусловием (while).

Описание функций. Вызов функции в выражении. Операторы циклов repeat, for.

Описание функций. Вызов функции в выражении. Оператор выбора (case).

Описание процедур. Операторы вызова процедуры, циклов repeat, for.

Описание процедур. Операторы вызова процедуры, выбора (case).

Пользовательские скалярные типы (перечислимый, интервальный). Описание

массивов. Индексированные переменные в выражении. Операторы циклов repeat, for.

Раздел описания констант. Пользовательские скалярные типы (перечислимый,

интервальный). Описание массивов. Индексированные переменные в выражении.

Описание записей (без вариантной части). Переменные - поля записи в выражении.

Операторы присоединения (with) , циклов repeat, for.

Описание записей (без вариантной части). Переменные - поля записи в выражении.

Операторы присоединения (with), выбора (case).

Интервальный тип. Описание массивов. Описание записей (без вариантной части).

Индексированные переменные и поля записи в выражении. Оператор присоединения

(with).

Интервальный тип. Описание массивов. Индексированные переменные в выражении.

Описание процедур. Оператор вызова процедуры.

Ссылочные типы данных. Описание функций. Вызов функции и указатели в

выражении.

Раздел описания констант. Ссылочные типы данных. Указатели в выражениях.

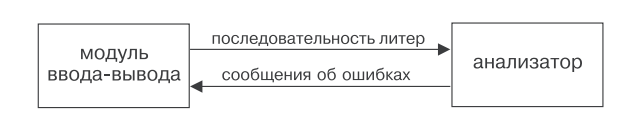
Оператор выбора (case).

Описание записей. Переменные - поля записи в выражении. Оператор присоединения

(with).

2 Модуль ввода-вывода

2.1 Описание

Модуль ввода-вывода считывает последовательность литер исходной программы с внешнего устройства и передает их анализатору. Анализатор проверяет, удовлетворяет ли эта последовательность литер правилам описания языка, и формирует(в случае необходимости) сообщения об ошибках. Такое взаимодействие между модулем ввода-вывода и анализатором можно представить в виде схемы: 

Для печати сообщений об ошибках будем хранить для каждого символа его позицию.

Сделаем ввод программы доступным в двух вариантах: по пути к файлу программы, либо же по строке в которой хранится код программы.

2.2 Проектирование

В модуле ввода-вывода должны быть определены следующие методы:

* Получение следующего символа
* Получение текущей позиции
* Генерация текста кода программы с указанием ошибок

Дополнительно для удобства создадим перечислимый тип OpenState – принимающий два состояния – открытие с помощью пути к файлу программы, либо же состояние, когда код программы был передан строкой.

Для поддержания текущей позиции заведем специальную переменную position.

В случае когда исходный код был получен с помощью файла, то текст кода с сгенерированными ошибками будем сохранять по тому же пути что и исходный, но с определенным суффиксом в конце имени файла. Для суффикса заведем строку-константу для удобства.

В функцию вывода ошибок передается указатель на экземпляр класса обработчика ошибок и вызывается соответствующая функция для получения всех ошибок.

После этого мы будем идти построчно, а затем посимвольно и сопоставлять текущую позицию с позициями ошибок(ошибки отсортированы по возрастанию позиции в которой находится ошибка).

Если в текущей позиции есть ошибка, то добавим ее в вектор ошибок текущей строки. После обработки всей строки дополнительно выведем информацию об ошибках и их позициях следующим образом:

Допустим у нас в строке нашлись следующие ошибки(позиции ошибок указаны относительно текущей строки): {2, ‘Вычисленное выражение имеет другой тип в отличие от переменной’}, {7, ‘Данную операцию нельзя применить к этим операндам’}

А текущая рассматриваемая строка такая: “*x := y + 1;*”

То текст с ошибками будет сгенерирован следующим образом:

*x := y + 1;*

*^ ^*

*| Данную операцию нельзя применить к этим операндам*

*Вычисленное выражение имеет другой тип в отличие от переменной*

2.3 Реализация

Ниже представлено описание класса и перечислимого типа в заголовочном файле IO\_Module.h

enum OpenState {

ByFile,

ByString

};

class IO\_Module

{

private:

const string OUTPUT\_PREFIX = "ANALYZED";

string path\_to\_file;

fstream input\_stream;

string code;

OpenState open\_state;

int position = -1;

string change\_filename(string& path);

public:

IO\_Module() {};

IO\_Module(const string& input);

char get\_next\_char();

void write\_errors(ErrorHandler\* error\_handler);

int get\_current\_position();

~IO\_Module();

};

Определим все методы в файле-источнике IO\_Module.cpp

IO\_Module::IO\_Module(const string& input)

{

try

{

input\_stream.open(input);

path\_to\_file = input;

open\_state = ByFile;

}

catch (exception e)

{

code = input;

open\_state = ByString;

}

}

IO\_Module::~IO\_Module()

{

input\_stream.close();

}

char IO\_Module::get\_next\_char()

{

position++;

char c;

if (input\_stream.is\_open())

{

c = (char)input\_stream.get();

}

else

{

if (position < code.size())

c = code[position];

else

c = EOF;

}

if (c == EOF)

input\_stream.close();

return c;

}

int IO\_Module::get\_current\_position()

{

return position;

}

void IO\_Module::write\_errors(ErrorHandler\* error\_handler)

{

vector<Error\*> errors = error\_handler->get\_errors();

IO\_Module\* io\_helper;

string result;

if (open\_state == ByFile)

io\_helper = new IO\_Module(path\_to\_file);

else

io\_helper = new IO\_Module(code);

int current\_error = 0;

char c;

do

{

// Построчно будем искать позиция с ошибками

int current\_position\_in\_line = 0;

ErrorHandler\* error\_handler\_in\_line = new ErrorHandler();

string first\_line;

do

{

c = io\_helper->get\_next\_char();

// Если еще есть ошибки и на текущей позиции есть ошибка, тогда добавим ее

if (current\_error < errors.size() &&

errors[current\_error]->position == io\_helper->get\_current\_position())

{

error\_handler\_in\_line->add\_error(errors[current\_error]->info, current\_position\_in\_line);

current\_error++;

}

current\_position\_in\_line++;

first\_line += (c == '**\t**' ? '**\t**' : ' ');

result += (c == EOF ? '**\n**' : c);

} while (c != EOF && c != '**\n**');

vector<Error\*> errors\_in\_line = error\_handler\_in\_line->get\_errors();

int errors\_count\_in\_line = errors\_in\_line.size();

if (errors\_count\_in\_line == 0) continue;

// Если в строке были ошибки то создадим первую строку типа:

// ^ ^ ^

// Указатели на позиции ошибок

for (auto e : errors\_in\_line)

first\_line[e->position] = '^';

first\_line += '**\n**';

result += first\_line;

for (int i = errors\_count\_in\_line - 1; i >= 0; i--)

{

// Обрежем первую строку до нужной позиции, т.к. на предыдущих итерациях мы уже обработали некоторые ошибки

string line = first\_line.substr(0, errors\_in\_line[i]->position);

// заменим на палочки, так красивее....

for (int j = 0; j < line.size(); j++)

{

if (line[j] == '^')

line[j] = '|';

}

// текущая ошибка всегда в конце строки, поэтому просто добавим описание ошибки в конец строки

line += errors\_in\_line[i]->info;

line += '**\n**';

result += line;

}

result += '**\n**';

delete error\_handler\_in\_line;

} while (c != EOF);

if (open\_state == ByFile)

{

string path\_to\_result = change\_filename(path\_to\_file);

ofstream fout(path\_to\_result);

fout << result;

fout.close();

}

else

{

cout << result;

}

delete io\_helper;

}

string IO\_Module::change\_filename(string& path)

{

// попытаемся найти где начинается расширение

string result;

auto pos = path.rfind(".");

if (pos == string::npos) // если у файла нет расширения

{

return path + OUTPUT\_PREFIX;

}

result = path.substr(0, pos) + "\_" + OUTPUT\_PREFIX + path.substr(pos);

return result;

}

2.4 Тестирование

2.4.1 Создание из строки исходного кода

const string PATH\_INPUT\_FILE = "C:**\\**Users**\\**user**\\**pascal.txt";

IO\_Module\* io = new IO\_Module(PATH\_INPUT\_FILE);

2.4.2 Создание из файла

const string SOURCE\_CODE = "program test; var x:integer; begin x:=5; end.";

IO\_Module\* io = new IO\_Module(SOURCE\_CODE);

2.4.3 Получение текущего символа, позиции

const string PATH\_INPUT\_FILE = "C:**\\**Users**\\**user**\\**pascal.txt";

IO\_Module\* io = new IO\_Module(PATH\_INPUT\_FILE);

cout << io->get\_next\_char() << " - " << io->get\_current\_position() << endl;

2.4.4 Вывод сообщения об ошибке

const string SOURCE\_CODE = "program test; var x:integer; begin y:=5; end.";

ErrorHandler\* error\_handler = new ErrorHandler();

IO\_Module\* io = new IO\_Module(SOURCE\_CODE);

string error\_text = "Переменная не была объявлена";

int error\_position = 35;

error\_handler->add\_error(error\_text, error\_position);

io->write\_errors(error\_handler);

Вывод: 

3 Лексический анализатор

3.1 Описание

Лексический анализатор формирует так называемые токены, объекты, созданные на основе последовательности символов(т.е. определенных слов заложенных в компиялторе).

Было выбрана реализация при которой лексический анализатор получает на вход объект модуля ввода-вывода и строит с помощью него последовательность этих токенов.

3.2 Проектирование

В модуле лексического анализатора были определены следующие методы:

* Получение токена
* Функция проверки программы – говорит о том были ли обнаружены какие-либо ошибки на этапе лексического анализа
* Функция получения всех токенов исходной программы

В сам конструктор класса нашего модуля должны быть переданы указатель на модуль ввода-вывода и указатель на обработчик ошибок.

Главной функцией безусловно является функция получения очередного токена, поэтому опишем ее более подробно.

Пропускаем все пробелы, знаки табуляции, переводы на новую строку, они нам не нужны, запоминаем текущую позицию – это будет позицией нашего будущего, еще несформированного токена.

Сделаем вывод каким будет токен исходя из его первого символа:

* Если символ – число, то токеном будет являться численная константа
* Если символ – буква, то токеном является какое-либо ключевое слово, либо идентификатор, либо это булевская константа(true/false). Для определения является ли слово ключевым, либо булевской константой будем использовать хэш-таблицы для быстрого поиска соответсвия. Если соответствия не нашлось – то слово является идентификатором.
* Если символ – ‘, то это должна быть константа типа Char.
* Если символ – “, то это должна быть константа типа String.
* Иначе это какой-либо оператор, например присваивания, сложения и т.п.

Будем продолжать генерировать и сохранять токены до тех пор, пока не дойдем до конца файла.

Токены разделим на три типа – токен-оператор, токен-идентификатор, токен-константа. Каждый токен должен хранить позицию начала данного токена.

Создадим абстрактный класс токен и будем наследоваться от него во всех дочерних классах токен-оператор, токен-идентификатор, токен-константа.

Класс токен-константа будет классом шаблоном и хранить значение константы, то есть использовать template, так как константы в нашем случае могут быть пяти типов – Int, Double, Bool, Char и String.

Класс токен-идентификатор будет хранить имя идентификатора.

Класс токен-оператор будет хранить тип оператора.

3.3 Реализация

3.3.1 Лексический анализатор

Ниже представлено описание класса в заголовочном файле LexicalAnalyzer.h

class LexicalAnalyzer

{

public:

IO\_Module\* io;

ErrorHandler\* error\_handler;

LexicalAnalyzer(IO\_Module\* \_io, ErrorHandler\* \_error\_handler);

bool check();

vector<Token\*> get\_tokens();

~LexicalAnalyzer();

private:

Token\* get\_token();

vector<Token\*> tokens;

int position = 0;

char c;

};

В файле LexicalAnalyzer.cpp представлены определения этих методов:

LexicalAnalyzer::LexicalAnalyzer(IO\_Module\* \_io, ErrorHandler\* \_error\_handler)

{

error\_handler = \_error\_handler;

io = \_io;

c = io->get\_next\_char();

}

Token\* LexicalAnalyzer::get\_token()

{

while (c == ' ' || c == '**\n**' || c == '**\r**' || c == '**\t**')

c = io->get\_next\_char();

position = io->get\_current\_position();

if (c == EOF)

{

return nullptr;

}

// Парсинг чисел

else if (isdigit(c))

{

string lexem(1, c);

c = io->get\_next\_char();

while (isdigit(c))

{

lexem += c;

c = io->get\_next\_char();

}

// если число слишком длинное - все плохо

if (c == '.') // типа real?

{

lexem += c;

c = io->get\_next\_char();

while (isdigit(c))

{

lexem += c;

c = io->get\_next\_char();

}

return new ConstToken<double>(ttConst, stod(lexem), dtReal, position);

}

else

{

return new ConstToken<int>(ttConst, stoi(lexem), dtInt, position);

}

}

// Парсинг идентификаторов/операторов

else if (isalpha(c))

{

string lexem(1, c);

c = io->get\_next\_char();

while (isdigit(c) || isalpha(c))

{

lexem += c;

c = io->get\_next\_char();

}

if (OperatorKeyWords.find(lexem) == OperatorKeyWords.end())

{

return new IdentificatorToken(ttIdentificator, lexem, position);

}

else

{

OperatorType ot = OperatorKeyWords.at(lexem);

if (ot == otTrue || ot == otFalse)

{

return new ConstToken<bool>(ttConst, ot, dtBool, position);

}

else

{

return new OperatorToken(ttOperator, ot, position);

}

}

}

// Парсинг символов

else if (c == '**\'**')

{

char lexem = io->get\_next\_char();

c = io->get\_next\_char();

if (c != '**\'**') // ошибка

{

string error\_text = "Ожидалася символ '";

error\_handler->add\_error(error\_text, io->get\_current\_position);

}

else

{

c = io->get\_next\_char();

return new ConstToken<char>(ttConst, lexem, dtChar, position);

}

}

// Парсинг строк

else if (c == '"')

{

string lexem = "";

c = io->get\_next\_char();

// если закрытия строки не будет - ошибку как-то

while (c != '"')

{

lexem += c;

c = io->get\_next\_char();

if (c == '**\n**' || c == EOF)

break;

}

c = io->get\_next\_char();

return new ConstToken<string>(ttConst, lexem, dtString, position);

}

// Парсинг небуквенных операторов

else

{

string lexem(1, c);

OperatorType ot = otError;

if (OperatorSymbols.find(lexem) != OperatorSymbols.end())

ot = OperatorSymbols.at(lexem);

*/\**

*.. | := | >= | <= | <>*

*\*/*

switch (ot)

{

case otDot:

c = io->get\_next\_char();

if (c == '.')

{

ot = otDots;

c = io->get\_next\_char();

}

break;

case otColon:

c = io->get\_next\_char();

if (c == '=')

{

ot = otAssign;

c = io->get\_next\_char();

}

break;

case otLess:

c = io->get\_next\_char();

if (c == '=')

{

ot = otLessEqual;

c = io->get\_next\_char();

}

else if (c == '>')

{

ot = otLessGreater;

c = io->get\_next\_char();

}

break;

case otGreater:

c = io->get\_next\_char();

if (c == '=')

{

ot = otGreaterEqual;

c = io->get\_next\_char();

}

break;

default:

c = io->get\_next\_char();

}

if (ot == otError)

return new Token(ttUndefined, position);

return new OperatorToken(ttOperator, ot, position);

}

}

bool LexicalAnalyzer::check()

{

int errors\_count = 0;

Token\* new\_token = get\_token();

do

{

tokens.push\_back(new\_token);

new\_token = get\_token();

} while (new\_token != nullptr);

tokens.push\_back(new Token(ttUndefined, io->get\_current\_position()));

return error\_handler->get\_errors\_count() == errors\_count;

}

vector<Token\*> LexicalAnalyzer::get\_tokens()

{

return tokens;

}

LexicalAnalyzer::~LexicalAnalyzer()

{

delete io;

}

3.3.2 Токен

enum TokenType {

ttIdentificator,

ttOperator,

ttConst,

ttUndefined

};

class Token {

public:

TokenType token\_type;

int position;

Token(TokenType token\_type, int position)

{

this->token\_type = token\_type;

this->position = position;

}

virtual ~Token() = default;

};

3.3.3 Токен-константа

template<typename T>

class ConstToken : public Token {

public:

T value;

DataType data\_type;

ConstToken(TokenType token\_type, T value, DataType data\_type, int position) : Token(token\_type, position)

{

this->value = value;

this->data\_type = data\_type;

}

};

3.3.4 Токен-идентификатор

class IdentificatorToken : public Token {

public:

string name;

IdentificatorToken(TokenType token\_type, string name, int position) : Token(token\_type, position)

{

this->name = name;

}

};

3.3.5 Токен-оператор

class OperatorToken : public Token {

public:

OperatorType operator\_type;

OperatorToken(TokenType token\_type, OperatorType operator\_type, int position) : Token(token\_type, position)

{

this->operator\_type = operator\_type;

}

};

3.3.6 Хэш-таблицы и перечислимый тип для операторов

enum OperatorType {

otError,

otPlus, // +

otMinus, // -

otSlash, // /

otStar, // \*

otDot, // .

otDots, // ..

otEqual, // =

otLess, // <

otGreater, // >

otAssign, // :=

otComma, // ,

...

otInteger, // integer

otBool, // bool

otReal, // real

otString, // string

otChar, // char

otTrue, // true

otFalse // false

};

const map<string, OperatorType> OperatorKeyWords = {

{"if", otIf},

{"do", otDo},

{"of", otOf},

{"or", otOr},

{"in", otIn},

{"to", otTo},

...

{"integer", otInteger},

{"bool", otBool},

{"real", otReal},

{"string", otString},

{"char", otChar},

{"true", otTrue},

{"false", otFalse}

};

const map<string, OperatorType> OperatorSymbols = {

{"+", otPlus},

{"-", otMinus},

{"/", otSlash},

{"\*", otStar},

{".", otDot},

{"..", otDots},

{"=", otEqual},

{"<", otLess},

{">", otGreater},

{":=", otAssign},

{",", otComma},

{">=", otGreaterEqual},

{"<=", otLessEqual},

{";", otSemiColon},

{":", otColon},

{"(", otLeftParenthesis},

{")", otRightParenthesis},

{"[", otLeftBracket},

{"]", otRightBracket},

{"<>", otLessGreater}

};

const map<OperatorType, string> KeyWordByOperator = {

{otIf, "if"},

{otDo, "do"},

{otOf, "of"},

{otOr, "or"},

{otIn, "in"},

{otTo, "to"},

...

{otLeftBracket, "["},

{otRightBracket, "]"},

{otLessGreater, "<>"},

{otTrue, "true"},

{otFalse, "false"}

};

3.4 Тестирование

Код программы:

**program** mfdsain;

**var** x:**integer**; y:**char**;

**begin**

**while** **not** (x > -5) **do**

**begin**

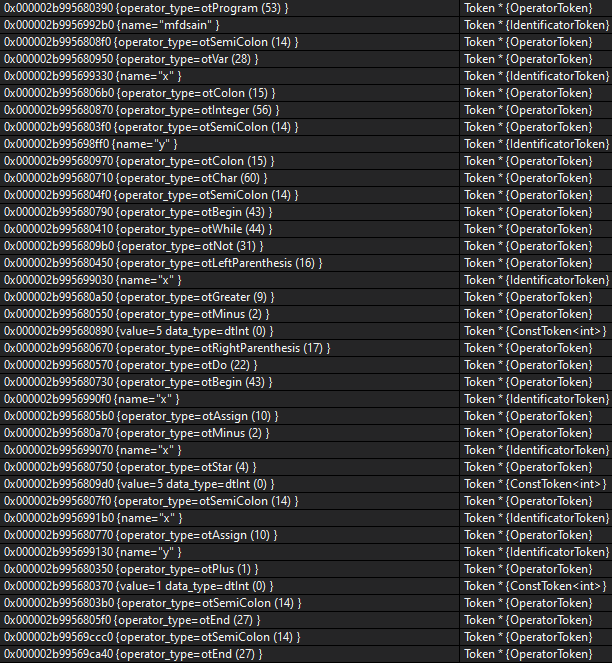
x := -x \* 5;

x := y + 1;

**end**;

**end**

Сгенерированные токены:



4 Синтаксический анализатор

4.1 Описание

Синтаксический анализатор проверяет, удовлетворяет ли программа формальным правилам.

Для задания синтаксиса широко применяются формальные правила, записанные в формах Бэкуса-Наура(БНФ), а также синтаксические диаграммы.

4.2 Проектирование

Для начала мы должны определиться какие формальные правила мы будем реализовывать, то есть нужно выписать все формы Бэкуса-Наура которые нам понадобятся:

<программа>::=program <имя>;<блок>.

<блок>::=<раздел переменных><раздел операторов>

<описание однотипных переменных>::=<имя>{,<имя>}:<тип>

<тип>::=integer|real|string|char

<раздел переменных>::= var <описание однотипных переменных>;{<описание однотипных переменных>;} | <пусто>

<раздел операторов>::= <составной оператор>

<оператор>::=<простой оператор>|<сложный оператор>

<простой оператор>::=<переменная>:=<выражение>

<переменная>::=<имя>

<выражение>::=<простое выражение>|<простое выражение><операция отношения><простое выражение>

<операция отношения>::= =|<>|<|<=|>=|>

<простое выражение>::=<слагаемое>{<аддитивная операция><слагаемое>}

<аддитивная операция>::= +|-|or

<слагаемое>::=<множитель>{<мультипликативная операция><множитель>}

<мультипликативная операция>::=\*|/|div|mod|and

<множитель>::=[<знак>]<переменная>|[<знак>]<константа>|[<знак>](<выражение>)|not <множитель>

<знак>::= +|-

<сложный оператор>::=<составной оператор>|<выбирающий оператор>|<оператор цикла>

<составной оператор>::= begin <оператор>{;<оператор>} end

<выбирающий оператор>::= if <выражение> then <оператор>|if <выражение> then <оператор> else <оператор>

<оператор цикла>::= while <выражение> do <оператор>

Для каждого описанного выше правила должны быть описана функция, тело которой есть результат преобразования правой части.

Наш синтаксический анализатор будет строиться по принципу детерминированного рекурсивного нисходящего алгоритма.

В модуле синтаксического анализатора были определены следующие методы:

* Функция пытающаяся “принять” очередной токен, то есть подается тип токена который может/должен идти следующим и эта функция проверяет выполняется ли это условие или нет. Эта же функция будет добавлять различные ошибки в обработчик ошибок.
* Функция переходящая на следующий токен
* Функция проверки программы – говорит о том является ли программа синтаксически правильной
* И все функции реализующие вышеописанные формы Бэкуса-Наура

Также были созданы различные переменные – вектор токенов, указатель на обработчик ошибок, указатель на текущий токен.

4.3 Реализация

4.3.1 Функция next\_token

void SyntaxAnalyzer::next\_token()

{

if (current\_token\_position == tokens.size())

return;

current\_token = tokens[current\_token\_position];

current\_token\_position++;

}

4.3.2 Функция accept

bool SyntaxAnalyzer::accept(TokenType token\_type, bool is\_necessarily)

{

bool result = true;

if (current\_token->token\_type != token\_type)

result = false;

if (result)

next\_token();

else if (is\_necessarily) // выводим ошибку

{

string error\_text = "";

switch (current\_token->token\_type)

{

case ttIdentificator:

error\_text = "Ожидалось имя идентификатора";

break;

case ttOperator:

error\_text = "Ожидалось имя оператора";

break;

case ttConst:

error\_text = "Ожидалась константа";

break;

}

error\_handler->add\_error(error\_text, current\_token->position);

next\_token();

}

return result;

}

bool SyntaxAnalyzer::accept(OperatorType operator\_type, bool is\_necessarily)

{

bool result = true;

if (current\_token->token\_type != ttOperator)

result = false;

if (((OperatorToken\*)current\_token)->operator\_type != operator\_type)

result = false;

if (result)

{

next\_token();

}

else if (is\_necessarily) // ошибочка вышла

{

string error\_text = "Ожидался оператор: " + KeyWordByOperator.find(operator\_type)->second;

error\_handler->add\_error(error\_text, current\_token->position);

next\_token();

}

return result;

}

bool SyntaxAnalyzer::accept(vector<OperatorType> operator\_types, bool is\_necessarily)

{

bool result = false;

if (current\_token->token\_type == ttOperator)

{

OperatorType current\_type = ((OperatorToken\*)current\_token)->operator\_type;

for (OperatorType operator\_type : operator\_types)

{

if (operator\_type == current\_type)

{

result = true;

break;

}

}

}

if (result)

{

next\_token();

}

else if (is\_necessarily)

{

string error\_text = "Ожидался один из операторов: ";

for (OperatorType operator\_type : operator\_types)

error\_text += KeyWordByOperator.find(operator\_type)->second + ", ";

error\_handler->add\_error(error\_text, current\_token->position);

next\_token();

}

return result;

}

4.3.3 Функции реализующие БНФ

void SyntaxAnalyzer::program() // <программа>::=program <имя>(<имя файла>{,<имя файла>});<блок>.

{

if (!accept(otProgram, true)) return;

if (!accept(ttIdentificator, true)) return;

if (!accept(otSemiColon, true)) return;

block();

if (!accept(otDot, true)) return;

}

void SyntaxAnalyzer::block() // <блок>::=<раздел констант><раздел типов><раздел переменных><раздел процедур и функций><раздел операторов>

{

//constants\_section();

vars\_section();

//functions\_section();

operators\_section();

}

// ======== Раздел переменных ========

bool SyntaxAnalyzer::single\_var\_definition() // <описание однотипных переменных>::=<имя>{,<имя>}:<тип>

{

if (!accept(ttIdentificator))

return false;

while (accept(otComma))

{

if (!accept(ttIdentificator, true)) return false;

}

if (!accept(otColon, true)) return false;

type();

return true;

}

void SyntaxAnalyzer::type() // <тип>::=integer|real|string|char

{

accept({ otInteger, otReal, otString, otChar }, true);

}

void SyntaxAnalyzer::vars\_section() // <раздел переменных>::= var <описание однотипных переменных>;{<описание однотипных переменных>;} | <пусто>

{

if (!accept(otVar))

return;

single\_var\_definition();

if (!accept(otSemiColon, true)) return;

while (single\_var\_definition())

{

if (!accept(otSemiColon, true)) return;

}

}

// ======== Раздел операторов ========

// <раздел операторов>::= <составной оператор>

void SyntaxAnalyzer::operators\_section()

{

neccessary\_compound\_operator();

}

//<оператор>::=<простой оператор>|<сложный оператор>

void SyntaxAnalyzer::operator\_()

{

if (!simple\_operator())

complex\_operator();

}

//<простой оператор>::=<переменная>:=<выражение>

bool SyntaxAnalyzer::simple\_operator() // \*

{

if (!var())

return false;

if (!accept(otAssign, true)) return false;

expression();

return true;

}

//<переменная>::=<имя>

bool SyntaxAnalyzer::var() // \*

{

return accept(ttIdentificator);

}

//<выражение>::=<простое выражение>|<простое выражение><операция отношения><простое выражение>

void SyntaxAnalyzer::expression()

{

simple\_expression();

if (relation\_operation())

simple\_expression();

}

//<операция отношения>::= =|<>|<|<=|>=|>

bool SyntaxAnalyzer::relation\_operation() // \*

{

return accept({ otEqual, otLessGreater, otLessEqual, otGreaterEqual, otGreater });

}

//<простое выражение>::=<слагаемое>{<аддитивная операция><слагаемое>}

void SyntaxAnalyzer::simple\_expression()

{

term();

while (additive\_operation())

term();

}

//<аддитивная операция>::= +|-|or

bool SyntaxAnalyzer::additive\_operation() // \*

{

return accept({ otPlus, otMinus, otOr });

}

//<слагаемое>::=<множитель>{<мультипликативная операция><множитель>}

void SyntaxAnalyzer::term()

{

factor();

while (multiplicative\_operation())

factor();

}

//<мультипликативная операция>::=\*|/|div|mod|and

bool SyntaxAnalyzer::multiplicative\_operation() // \*

{

return accept({ otStar, otSlash, otDiv, otMod, otAnd });

}

//<множитель>::=[<знак>]<переменная>|[<знак>]<константа>|[<знак>](<выражение>)|not <множитель>

void SyntaxAnalyzer::factor()

{

if (sign())

{

//...

}

if (var())

{

//...

}

else if (accept(ttConst))

{

//...

}

else if (accept(otLeftParenthesis))

{

expression();

accept(otRightParenthesis, true);

}

else if (accept(otNot))

{

factor();

}

else

{

string error\_text = "Ожидалась переменная/константа/выражение";

error\_handler->add\_error(error\_text, current\_token->position);

}

}

//<знак>::= +|-

bool SyntaxAnalyzer::sign() // \*

{

return accept({ otPlus, otMinus });

}

//<сложный оператор>::=<составной оператор>|<выбирающий оператор>|<оператор цикла>

void SyntaxAnalyzer::complex\_operator()

{

if (compound\_operator())

{

//...

}

else if (if\_operator())

{

//...

}

else if (while\_operator())

{

//...

}

}

//<составной оператор>::= begin <оператор>{;<оператор>} end

bool SyntaxAnalyzer::compound\_operator() // \*

{

if (!accept(otBegin))

return false;

operator\_();

while (accept(otSemiColon))

{

operator\_();

}

if (!accept(otEnd, true)) return false;

return true;

}

//<обязательный составной оператор>::= begin <оператор>{;<оператор>} end

void SyntaxAnalyzer::neccessary\_compound\_operator() // \*

{

if (!accept(otBegin, true)) return;

operator\_();

while (accept(otSemiColon))

{

operator\_();

}

if (!accept(otEnd, true)) return;

}

//<выбирающий оператор>::= if <выражение> then <оператор>|if <выражение> then <оператор> else <оператор>

bool SyntaxAnalyzer::if\_operator() // \*

{

if (!accept(otIf))

return false;

expression();

accept(otThen);

operator\_();

if (accept(otElse))

operator\_();

return true;

}

//<оператор цикла>::= while <выражение> do <оператор>

bool SyntaxAnalyzer::while\_operator() // \*

{

if (!accept(otWhile))

return false;

expression();

if (!accept(otDo, true)) return false;

operator\_();

return true;

}

4.3.4 Полное описание класса

Ниже представлено описание класса в заголовочном файле SyntaxAnalyzer.h

class SyntaxAnalyzer

{

public:

SyntaxAnalyzer(vector<Token\*> \_tokens, ErrorHandler\* \_error\_handler);

bool check();

~SyntaxAnalyzer();

private:

vector<Token\*> tokens;

ErrorHandler\* error\_handler;

int current\_token\_position;

Token\* current\_token;

void next\_token();

bool accept(TokenType token\_type, bool is\_necessarily = false);

bool accept(OperatorType operator\_type, bool is\_necessarily = false);

bool accept(vector<OperatorType> operator\_types, bool is\_necessarily = false);

void program();

void block();

void vars\_section();

void operators\_section();

bool single\_var\_definition();

void type();

void operator\_();

bool simple\_operator();

bool var();

void expression();

bool relation\_operation();

void simple\_expression();

bool additive\_operation();

void term();

bool multiplicative\_operation();

void factor();

bool sign();

void complex\_operator();

void neccessary\_compound\_operator();

bool compound\_operator();

bool if\_operator();

bool while\_operator();

};

Определим все методы в файле-источнике SyntaxAnalyzer.cpp

SyntaxAnalyzer::SyntaxAnalyzer(vector<Token\*> \_tokens, ErrorHandler\* \_error\_handler)

{

error\_handler = \_error\_handler;

tokens = \_tokens;

current\_token\_position = 0;

next\_token();

}

SyntaxAnalyzer::~SyntaxAnalyzer()

{

delete current\_token;

}

bool SyntaxAnalyzer::check()

{

int errros\_count = error\_handler->get\_errors\_count();

program();

return error\_handler->get\_errors\_count() == errros\_count;

}

void SyntaxAnalyzer::next\_token()

{

if (current\_token\_position == tokens.size())

return;

current\_token = tokens[current\_token\_position];

current\_token\_position++;

}

bool SyntaxAnalyzer::accept(TokenType token\_type, bool is\_necessarily)

{

bool result = true;

if (current\_token->token\_type != token\_type)

result = false;

if (result)

next\_token();

else if (is\_necessarily) // выводим ошибку

{

string error\_text = "";

switch (current\_token->token\_type)

{

case ttIdentificator:

error\_text = "Ожидалось имя идентификатора";

break;

case ttOperator:

error\_text = "Ожидалось имя оператора";

break;

case ttConst:

error\_text = "Ожидалась константа";

break;

}

error\_handler->add\_error(error\_text, current\_token->position);

next\_token();

}

return result;

}

bool SyntaxAnalyzer::accept(OperatorType operator\_type, bool is\_necessarily)

{

bool result = true;

if (current\_token->token\_type != ttOperator)

result = false;

if (((OperatorToken\*)current\_token)->operator\_type != operator\_type)

result = false;

if (result)

{

next\_token();

}

else if (is\_necessarily) // ошибочка вышла

{

string error\_text = "Ожидался оператор: " + KeyWordByOperator.find(operator\_type)->second;

error\_handler->add\_error(error\_text, current\_token->position);

next\_token();

}

return result;

}

bool SyntaxAnalyzer::accept(vector<OperatorType> operator\_types, bool is\_necessarily)

{

bool result = false;

if (current\_token->token\_type == ttOperator)

{

OperatorType current\_type = ((OperatorToken\*)current\_token)->operator\_type;

for (OperatorType operator\_type : operator\_types)

{

if (operator\_type == current\_type)

{

result = true;

break;

}

}

}

if (result)

{

next\_token();

}

else if (is\_necessarily)

{

string error\_text = "Ожидался один из операторов: ";

for (OperatorType operator\_type : operator\_types)

error\_text += KeyWordByOperator.find(operator\_type)->second + ", ";

error\_handler->add\_error(error\_text, current\_token->position);

next\_token();

}

return result;

}

void SyntaxAnalyzer::program() // <программа>::=program <имя>(<имя файла>{,<имя файла>});<блок>.

{

if (!accept(otProgram, true)) return;

if (!accept(ttIdentificator, true)) return;

if (!accept(otSemiColon, true)) return;

block();

if (!accept(otDot, true)) return;

}

void SyntaxAnalyzer::block() // <блок>::=<раздел констант><раздел типов><раздел переменных><раздел процедур и функций><раздел операторов>

{

//constants\_section();

vars\_section();

//functions\_section();

operators\_section();

}

// ======== Раздел переменных ========

bool SyntaxAnalyzer::single\_var\_definition() // <описание однотипных переменных>::=<имя>{,<имя>}:<тип>

{

if (!accept(ttIdentificator))

return false;

while (accept(otComma))

{

if (!accept(ttIdentificator, true)) return false;

}

if (!accept(otColon, true)) return false;

type();

return true;

}

void SyntaxAnalyzer::type() // <тип>::=integer|real|string|char

{

accept({ otInteger, otReal, otString, otChar }, true);

}

void SyntaxAnalyzer::vars\_section() // <раздел переменных>::= var <описание однотипных переменных>;{<описание однотипных переменных>;} | <пусто>

{

if (!accept(otVar))

return;

single\_var\_definition();

if (!accept(otSemiColon, true)) return;

while (single\_var\_definition())

{

if (!accept(otSemiColon, true)) return;

}

}

// ======== Раздел операторов ========

// <раздел операторов>::= <составной оператор>

void SyntaxAnalyzer::operators\_section()

{

neccessary\_compound\_operator();

}

//<оператор>::=<простой оператор>|<сложный оператор>

void SyntaxAnalyzer::operator\_()

{

if (!simple\_operator())

complex\_operator();

}

//<простой оператор>::=<переменная>:=<выражение>

bool SyntaxAnalyzer::simple\_operator() // \*

{

if (!var())

return false;

if (!accept(otAssign, true)) return false;

expression();

return true;

}

//<переменная>::=<имя>

bool SyntaxAnalyzer::var() // \*

{

return accept(ttIdentificator);

}

//<выражение>::=<простое выражение>|<простое выражение><операция отношения><простое выражение>

void SyntaxAnalyzer::expression()

{

simple\_expression();

if (relation\_operation())

simple\_expression();

}

//<операция отношения>::= =|<>|<|<=|>=|>

bool SyntaxAnalyzer::relation\_operation() // \*

{

return accept({ otEqual, otLessGreater, otLessEqual, otGreaterEqual, otGreater });

}

//<простое выражение>::=<слагаемое>{<аддитивная операция><слагаемое>}

void SyntaxAnalyzer::simple\_expression()

{

term();

while (additive\_operation())

term();

}

//<аддитивная операция>::= +|-|or

bool SyntaxAnalyzer::additive\_operation() // \*

{

return accept({ otPlus, otMinus, otOr });

}

//<слагаемое>::=<множитель>{<мультипликативная операция><множитель>}

void SyntaxAnalyzer::term()

{

factor();

while (multiplicative\_operation())

factor();

}

//<мультипликативная операция>::=\*|/|div|mod|and

bool SyntaxAnalyzer::multiplicative\_operation() // \*

{

return accept({ otStar, otSlash, otDiv, otMod, otAnd });

}

//<множитель>::=[<знак>]<переменная>|[<знак>]<константа>|[<знак>](<выражение>)|not <множитель>

void SyntaxAnalyzer::factor()

{

if (sign())

{

//...

}

if (var())

{

//...

}

else if (accept(ttConst))

{

//...

}

else if (accept(otLeftParenthesis))

{

expression();

accept(otRightParenthesis, true);

}

else if (accept(otNot))

{

factor();

}

else

{

string error\_text = "Ожидалась переменная/константа/выражение";

error\_handler->add\_error(error\_text, current\_token->position);

}

}

//<знак>::= +|-

bool SyntaxAnalyzer::sign() // \*

{

return accept({ otPlus, otMinus });

}

//<сложный оператор>::=<составной оператор>|<выбирающий оператор>|<оператор цикла>

void SyntaxAnalyzer::complex\_operator()

{

if (compound\_operator())

{

//...

}

else if (if\_operator())

{

//...

}

else if (while\_operator())

{

//...

}

}

//<составной оператор>::= begin <оператор>{;<оператор>} end

bool SyntaxAnalyzer::compound\_operator() // \*

{

if (!accept(otBegin))

return false;

operator\_();

while (accept(otSemiColon))

{

operator\_();

}

if (!accept(otEnd, true)) return false;

return true;

}

//<обязательный составной оператор>::= begin <оператор>{;<оператор>} end

void SyntaxAnalyzer::neccessary\_compound\_operator() // \*

{

if (!accept(otBegin, true)) return;

operator\_();

while (accept(otSemiColon))

{

operator\_();

}

if (!accept(otEnd, true)) return;

}

//<выбирающий оператор>::= if <выражение> then <оператор>|if <выражение> then <оператор> else <оператор>

bool SyntaxAnalyzer::if\_operator() // \*

{

if (!accept(otIf))

return false;

expression();

accept(otThen);

operator\_();

if (accept(otElse))

operator\_();

return true;

}

//<оператор цикла>::= while <выражение> do <оператор>

bool SyntaxAnalyzer::while\_operator() // \*

{

if (!accept(otWhile))

return false;

expression();

if (!accept(otDo, true)) return false;

operator\_();

return true;

}

4.4 Тестирование

Код программы:

**program** mfdsain;

**var** x:**integer**; y:**char**;

**begin**

**while** **not** (x > -5) **do**

**begin**

x := -x \* 5;

x := x + 1;

**end**;

**end**

Вывод:

**program** mfdsain;

**var** x:**integer**; y:**char**;

**begin**

**while** **not** (x > -5) **do**

**begin**

x := -x \* 5;

x := x + 1;

**end**;

**end**

^

Ожидался оператор: .

Код программы:

**program** mfdsain;

**var** x:**integer**; y:**char**;

**begin**

**not** (x > -5) **do**

**begin**

x := -x \* 5;

x := x + 1;

**end**;

**end**

Вывод:

**program** mfdsain;

**var** x:**integer**; y:**char**;

**begin**

**not** (x > -5) **do**

^ ^

| Ожидался оператор: .

Ожидался оператор: **end**

**begin**

x := -x \* 5;

x := x + 1;

**end**;

**end**

5 Семантический анализатор

5.1 Описание

Семантический анализатор получает на вход синтаксически проверенную программу и проверяет не нарушаются ли неформальные правила описания языка.

Реализованные проверки неформальных правил:

* Переменная может быть только одного типа
* Использование необъявленных переменных
* Использование переменных с неприводимыми типами в одном выражении
* Попытка присвоить значение выражения другого типа в отличие от переменной
* Использование неприводимого выражения к типу Boolean в выбирающем операторе и операторе цикла

5.2 Проектирование

Для типов были спроектированы и реализованы специальные классы, для каждого типа свой класс. В каждом таком классе была определена функция – приводим ли этот тип к другому типу?

Для этого был написан абстрактный класс Type и дочерние классы – IntegerType, BoolType, RealType, CharType, StringType

В модуле семантического анализатора были созданы следующие хэш-таблицы:

* Переменные {название, указатель на экземпляр типа}
* Доступные типы {тип данных(перечислимый тип из модуля с токенами), указатель на экземпляр типа}

При вызове конструктора добавим все пять типов в хэш-таблицу доступных типов.

При попытке добавления новой переменной сначала проверим есть ли переменная с таким именем в таблице, если есть – ошибка, иначе добавим эту пару <переменная, тип> в мэпу.

Также одной из важнейших функций является функция приведения типов – она получает два типа и пытается их привести к какому-то, например: [Integer, Real] -> можно привести к типу Real, этот тип и вернет функция.

5.3 Реализация

5.3.1 Функция добавления переменной

void SemanticAnalyzer::add\_var(VarName name, Type\* dt)

{

if (variables.find(name) == variables.end())

variables[name] = dt;

else

{

string error\_text = "Переменная с именем `" + name + "` уже была объявлена";

error\_handler->add\_error(error\_text, current\_token->position);

}

}

5.3.2 Функция приведения типов

Type\* SemanticAnalyzer::derive(Type\* left, Type\* right, OperatorType last\_operation, int position\_for\_error)

{

Type\* result = new Type();

if (left->can\_cast\_to(right))

result = right;

if (right->can\_cast\_to(left))

result = left;

bool is\_string = result->can\_cast\_to(available\_types[dtString]);

*/\**

*=|<>|<|<=|>=|>*

*accept({ otEqual, otLessGreater, otLessEqual, otGreaterEqual, otGreater });*

*+|-|or*

*accept({ otPlus, otMinus, otOr });*

*\*|/|div|mod|and*

*accept({ otStar, otSlash, otDiv, otMod, otAnd });*

*\*/*

string error\_text;

if (is\_string)

{

// только +, =, <>

if (last\_operation != otPlus && last\_operation != otEqual && last\_operation != otLessGreater)

{

error\_text = "Данную операцию нельзя применить к этим операндам";

error\_handler->add\_error(error\_text, position\_for\_error);

}

}

else

{

if (left->can\_cast\_to(available\_types[dtString])|| right->can\_cast\_to(available\_types[dtString])) // хотя бы один операнд - строковый

{

error\_text = "Данную операцию нельзя применить к этим операндам";

error\_handler->add\_error(error\_text, position\_for\_error);

}

}

if (last\_operation == otEqual || last\_operation == otLessGreater || last\_operation == otLessEqual ||

last\_operation == otGreaterEqual || last\_operation == otGreater || last\_operation == otLess ||

last\_operation == otOr || last\_operation == otAnd)

return available\_types[dtBool];

return result;

}

5.3.3 Полное описание класса модуля семантического анализатора

Описание в заголовочном файле SemanticAnalyzer.h

class SemanticAnalyzer

{

public:

SemanticAnalyzer(vector<Token\*> \_tokens, ErrorHandler\* \_error\_handler);

~SemanticAnalyzer();

bool check();

private:

int current\_token\_position;

Token\* current\_token;

vector<Token\*> tokens;

ErrorHandler\* error\_handler;

map<VarName, Type\*> variables;

map<DataType, Type\*> available\_types;

map<OperatorType, int> get\_last\_position\_of\_operator;

OperatorType lastOp;

void next\_token();

Type\* derive(Type\* left, Type\* right, OperatorType last\_operation, int position\_for\_error);

void add\_var(VarName name, Type\* dt);

VarName get\_var\_name\_from\_token(Token\* token);

Type\* get\_type\_from\_const\_token(Token\* token);

bool accept(TokenType token\_type);

bool accept(OperatorType operator\_type);

bool accept(vector<OperatorType> operator\_types);

void program();

void block();

void vars\_section();

void operators\_section();

bool single\_var\_definition();

Type\* type();

void operator\_();

bool simple\_operator();

Type\* expression();

bool relation\_operation();

Type\* simple\_expression();

bool additive\_operation();

Type\* term();

bool multiplicative\_operation();

Type\* factor();

bool sign();

void complex\_operator();

void neccessary\_compound\_operator();

bool compound\_operator();

bool if\_operator();

bool while\_operator();

};

Реализация описанных методов в файле-источнике SemanticAnalyzer.cpp:

SemanticAnalyzer::SemanticAnalyzer(vector<Token\*> \_tokens, ErrorHandler\* \_error\_handler)

{

tokens = \_tokens;

error\_handler = \_error\_handler;

current\_token\_position = 0;

next\_token();

available\_types[dtInt] = new IntegerType();

available\_types[dtReal] = new RealType();

available\_types[dtString] = new StringType();

available\_types[dtBool] = new BoolType();

available\_types[dtChar] = new CharType();

}

SemanticAnalyzer::~SemanticAnalyzer()

{

delete error\_handler;

}

bool SemanticAnalyzer::check()

{

int errors\_count = error\_handler->get\_errors\_count();

program();

return error\_handler->get\_errors\_count() == errors\_count;

}

void SemanticAnalyzer::next\_token()

{

if (current\_token\_position == tokens.size())

return;

current\_token = tokens[current\_token\_position];

current\_token\_position++;

if (current\_token->token\_type == ttOperator)

get\_last\_position\_of\_operator[((OperatorToken\*)current\_token)->operator\_type] = current\_token->position;

}

Type\* SemanticAnalyzer::derive(Type\* left, Type\* right, OperatorType last\_operation, int position\_for\_error)

{

Type\* result = new Type();

if (left->can\_cast\_to(right))

result = right;

if (right->can\_cast\_to(left))

result = left;

bool is\_string = result->can\_cast\_to(available\_types[dtString]);

*/\**

*=|<>|<|<=|>=|>*

*accept({ otEqual, otLessGreater, otLessEqual, otGreaterEqual, otGreater });*

*+|-|or*

*accept({ otPlus, otMinus, otOr });*

*\*|/|div|mod|and*

*accept({ otStar, otSlash, otDiv, otMod, otAnd });*

*\*/*

string error\_text;

if (is\_string)

{

// только +, =, <>

if (last\_operation != otPlus && last\_operation != otEqual && last\_operation != otLessGreater)

{

error\_text = "Данную операцию нельзя применить к этим операндам";

error\_handler->add\_error(error\_text, position\_for\_error);

}

}

else

{

if (left->can\_cast\_to(available\_types[dtString])|| right->can\_cast\_to(available\_types[dtString])) // хотя бы один операнд - строковый

{

error\_text = "Данную операцию нельзя применить к этим операндам";

error\_handler->add\_error(error\_text, position\_for\_error);

}

}

if (last\_operation == otEqual || last\_operation == otLessGreater || last\_operation == otLessEqual ||

last\_operation == otGreaterEqual || last\_operation == otGreater || last\_operation == otLess ||

last\_operation == otOr || last\_operation == otAnd)

return available\_types[dtBool];

return result;

}

void SemanticAnalyzer::add\_var(VarName name, Type\* dt)

{

if (variables.find(name) == variables.end())

variables[name] = dt;

else

{

string error\_text = "Переменная с именем `" + name + "` уже была объявлена";

error\_handler->add\_error(error\_text, current\_token->position);

}

}

VarName SemanticAnalyzer::get\_var\_name\_from\_token(Token\* token)

{

if (token->token\_type != ttIdentificator)

return "";

return ((IdentificatorToken\*)token)->name;

}

Type\* SemanticAnalyzer::get\_type\_from\_const\_token(Token\* token)

{

if (auto ct = dynamic\_cast<ConstToken<int>\*>(token)) {

return available\_types[ct->data\_type];

}

else if (auto ct = dynamic\_cast<ConstToken<double>\*>(token)) {

return available\_types[ct->data\_type];

}

else if (auto ct = dynamic\_cast<ConstToken<string>\*>(token)) {

return available\_types[ct->data\_type];

}

else if (auto ct = dynamic\_cast<ConstToken<char>\*>(token)) {

return available\_types[ct->data\_type];

}

else if (auto ct = dynamic\_cast<ConstToken<bool>\*>(token)) {

return available\_types[ct->data\_type];

}

return new Type();

}

bool SemanticAnalyzer::accept(TokenType token\_type)

{

bool result = true;

if (current\_token->token\_type != token\_type)

result = false;

if (result)

next\_token();

return result;

}

bool SemanticAnalyzer::accept(OperatorType operator\_type)

{

bool result = true;

if (current\_token->token\_type != ttOperator)

result = false;

if (((OperatorToken\*)current\_token)->operator\_type != operator\_type)

result = false;

if (result)

{

lastOp = operator\_type;

next\_token();

}

return result;

}

bool SemanticAnalyzer::accept(vector<OperatorType> operator\_types)

{

bool result = false;

if (current\_token->token\_type == ttOperator)

{

OperatorType current\_type = ((OperatorToken\*)current\_token)->operator\_type;

for (OperatorType operator\_type : operator\_types)

{

if (operator\_type == current\_type)

{

lastOp = operator\_type;

result = true;

break;

}

}

}

if (result)

next\_token();

return result;

}

void SemanticAnalyzer::program() // <программа>::=program <имя>(<имя файла>{,<имя файла>});<блок>.

{

accept(otProgram);

accept(ttIdentificator);

accept(otSemiColon);

block();

accept(otDot);

}

void SemanticAnalyzer::block() // <блок>::=<раздел констант><раздел типов><раздел переменных><раздел процедур и функций><раздел операторов>

{

vars\_section();

operators\_section();

}

// ======== Раздел переменных ========

bool SemanticAnalyzer::single\_var\_definition() // <описание однотипных переменных>::=<имя>{,<имя>}:<тип>

{

vector<VarName> variableNames;

variableNames.push\_back(get\_var\_name\_from\_token(current\_token));

if (!accept(ttIdentificator))

return false;

while (accept(otComma))

{

variableNames.push\_back(get\_var\_name\_from\_token(current\_token));

accept(ttIdentificator);

}

accept(otColon);

Type\* varType = type();

for (VarName name : variableNames)

{

add\_var(name, varType);

}

return true;

}

Type\* SemanticAnalyzer::type() // <тип>::=integer|real|string|char

{

if (accept(otInteger))

return available\_types[dtInt];

if (accept(otReal))

return available\_types[dtReal];

if (accept(otString))

return available\_types[dtString];

if (accept(otChar))

return available\_types[dtChar];

}

void SemanticAnalyzer::vars\_section() // <раздел переменных>::= var <описание однотипных переменных>;{<описание однотипных переменных>;} | <пусто>

{

accept(otVar);

single\_var\_definition();

accept(otSemiColon);

while (single\_var\_definition())

{

accept(otSemiColon);

}

}

// ======== Раздел операторов ========

// <раздел операторов>::= <составной оператор>

void SemanticAnalyzer::operators\_section()

{

neccessary\_compound\_operator();

}

//<оператор>::=<простой оператор>|<сложный оператор>

void SemanticAnalyzer::operator\_()

{

if (!simple\_operator())

complex\_operator();

}

//<простой оператор>::=<переменная>:=<выражение>

bool SemanticAnalyzer::simple\_operator() // \*

{

int mem\_position = current\_token->position;

VarName name = get\_var\_name\_from\_token(current\_token);

if (!accept(ttIdentificator))

return false;

accept(otAssign);

Type\* t = expression();

if (variables.find(name) == variables.end())

{

string error\_text = "Переменная не была объявлена";

error\_handler->add\_error(error\_text, mem\_position);

}

else

{

if (!t->can\_cast\_to(variables[name]))

{

// TODO: вывод ошибки

string error\_text = "Вычисленное выражение имеет другой тип в отличие от переменной";

error\_handler->add\_error(error\_text, get\_last\_position\_of\_operator[otAssign]);

}

}

return true;

}

//<выражение>::=<простое выражение>|<простое выражение><операция отношения><простое выражение>

Type\* SemanticAnalyzer::expression()

{

Type \*t1, \*t2;

t1 = simple\_expression();

if (relation\_operation())

{

OperatorType last\_operation = lastOp;

int position = get\_last\_position\_of\_operator[last\_operation];

t2 = simple\_expression();

t1 = derive(t1, t2, last\_operation, position);

}

return t1;

}

//<операция отношения>::= =|<>|<|<=|>=|>

bool SemanticAnalyzer::relation\_operation() // \*

{

return accept({ otEqual, otLessGreater, otLessEqual, otGreaterEqual, otGreater });

}

//<простое выражение>::=<слагаемое>{<аддитивная операция><слагаемое>}

Type\* SemanticAnalyzer::simple\_expression()

{

Type\* t1, \* t2;

t1 = term();

while (additive\_operation())

{

OperatorType last\_operation = lastOp;

int position = get\_last\_position\_of\_operator[last\_operation];

t2 = term();

t1 = derive(t1, t2, last\_operation, position);

}

return t1;

}

//<аддитивная операция>::= +|-|or

bool SemanticAnalyzer::additive\_operation() // \*

{

return accept({ otPlus, otMinus, otOr });

}

//<слагаемое>::=<множитель>{<мультипликативная операция><множитель>}

Type\* SemanticAnalyzer::term()

{

Type\* t1, \* t2;

t1 = factor();

while (multiplicative\_operation())

{

OperatorType last\_operation = lastOp;

int position = get\_last\_position\_of\_operator[last\_operation];

t2 = factor();

t1 = derive(t1, t2, last\_operation, position);

}

return t1;

}

//<мультипликативная операция>::=\*|/|div|mod|and

bool SemanticAnalyzer::multiplicative\_operation() // \*

{

return accept({ otStar, otSlash, otDiv, otMod, otAnd });

}

//<множитель>::=[<знак>]<переменная>|[<знак>]<константа>|[<знак>](<выражение>)|not <множитель>

Type\* SemanticAnalyzer::factor()

{

// TODO: сделать что-то со знаком...

if (sign())

{

//...

}

VarName name = get\_var\_name\_from\_token(current\_token);

Type\* const\_type = get\_type\_from\_const\_token(current\_token);

int mem\_position = current\_token->position;

if (accept(ttIdentificator))

{

if (variables.find(name) == variables.end())

{

string error\_text = "Переменная не была объявлена";

error\_handler->add\_error(error\_text, mem\_position);

return new Type();

}

return variables[name];

}

else if (accept(ttConst))

{

return const\_type;

}

else if (accept(otLeftParenthesis))

{

Type\* t = expression();

accept(otRightParenthesis);

return t;

}

else if (accept(otNot))

{

Type\* t = factor();

if (!t->can\_cast\_to(available\_types[dtBool]))

{

string error\_text = "Выражение должно иметь тип Bool";

error\_handler->add\_error(error\_text, get\_last\_position\_of\_operator[otNot]);

// TODO: ошибка

}

return available\_types[dtBool];

}

}

//<знак>::= +|-

bool SemanticAnalyzer::sign() // \*

{

return accept({ otPlus, otMinus });

}

//<сложный оператор>::=<составной оператор>|<выбирающий оператор>|<оператор цикла>

void SemanticAnalyzer::complex\_operator()

{

if (compound\_operator())

{

//...

}

else if (if\_operator())

{

//...

}

else if (while\_operator())

{

//...

}

}

//<составной оператор>::= begin <оператор>{;<оператор>} end

bool SemanticAnalyzer::compound\_operator() // \*

{

if (!accept(otBegin))

return false;

operator\_();

while (accept(otSemiColon))

{

operator\_();

}

accept(otEnd);

return true;

}

//<обязательный составной оператор>::= begin <оператор>{;<оператор>} end

void SemanticAnalyzer::neccessary\_compound\_operator() // \*

{

accept(otBegin);

operator\_();

while (accept(otSemiColon))

{

operator\_();

}

accept(otEnd);

}

//<выбирающий оператор>::= if <выражение> then <оператор>|if <выражение> then <оператор> else <оператор>

bool SemanticAnalyzer::if\_operator() // \*

{

if (!accept(otIf))

return false;

Type\* t = expression();

if (!t->can\_cast\_to(available\_types[dtBool]))

{

string error\_text = "Выражение должно иметь тип Bool";

error\_handler->add\_error(error\_text, current\_token->position);

// TODO: ошибка

}

accept(otThen);

operator\_();

if (accept(otElse))

operator\_();

return true;

}

//<оператор цикла>::= while <выражение> do <оператор>

bool SemanticAnalyzer::while\_operator() // \*

{

if (!accept(otWhile))

return false;

Type\* t = expression();

if (!t->can\_cast\_to(available\_types[dtBool]))

{

string error\_text = "Выражение должно иметь тип Bool";

error\_handler->add\_error(error\_text, current\_token->position);

// TODO: ошибка

}

accept(otDo);

operator\_();

return true;

}

5.3.4 Описание класса типов

enum EType

{

et\_intger,

et\_bool,

et\_real,

et\_char,

et\_string,

et\_undefined

};

class Type

{

public:

EType type = et\_undefined;

Type() {};

virtual bool can\_cast\_to(Type\* another\_type) { return false; };

};

class IntegerType : public Type

{

public:

IntegerType() { type = et\_intger; }

bool can\_cast\_to(Type\* another\_type)

{

return another\_type->type == et\_real ||

another\_type->type == type;

}

};

class BoolType : public Type

{

public:

BoolType() { type = et\_bool; }

bool can\_cast\_to(Type\* another\_type)

{

return another\_type->type == et\_real ||

another\_type->type == et\_intger ||

another\_type->type == et\_real ||

another\_type->type == type;

}

};

class RealType : public Type

{

public:

RealType() { type = et\_real; }

bool can\_cast\_to(Type\* another\_type)

{

return another\_type->type == type;

}

};

class CharType : public Type

{

public:

CharType() { type = et\_char; }

bool can\_cast\_to(Type\* another\_type)

{

return another\_type->type == et\_string || another\_type->type == type;

}

};

class StringType : public Type

{

public:

StringType() { type = et\_string; }

bool can\_cast\_to(Type\* another\_type)

{

return another\_type->type == type;

}

};

5.4 Тестирование

Код программы:

**program** mfdsain;

**var** x:**integer**; y:**char**;

**begin**

**while** **not** (x > -5) **do**

**begin**

**if** (1 > (x + (test + (1 + 'c')))) **then**

**begin**

x := -x \* 5;

x := y + 1;

f := 5

**end**;

x := -x \* 5;

x := y + 1;

f := 5

**end**;

**if** (1 > (x + (test + (1 + 'c')))) **then**

**begin**

z := 5

**end**;

**end**.

Вывод:

**program** mfdsain;

**var** x:**integer**; y:**char**;

**begin**

**while** **not** (x > -5) **do**

**begin**

**if** (1 > (x + (test + (1 + 'c')))) **then**

^ ^

| Данную операцию нельзя применить к этим операндам

Переменная не была объявлена

**begin**

x := -x \* 5;

x := y + 1;

^ ^

| Данную операцию нельзя применить к этим операндам

Вычисленное выражение имеет другой тип в отличие от переменной

f := 5

^

Переменная не была объявлена

**end**;

x := -x \* 5;

x := y + 1;

^ ^

| Данную операцию нельзя применить к этим операндам

Вычисленное выражение имеет другой тип в отличие от переменной

f := 5

^

Переменная не была объявлена

**end**;

**if** (1 > (x + (test + (1 + 'c')))) **then**

^ ^

| Данную операцию нельзя применить к этим операндам

Переменная не была объявлена

**begin**

z := 5

^

Переменная не была объявлена

**end**;

**end**.

Код программы:

**program** mfdsain;

**var** x:**integer**; y:**char**;

**begin**

**if** (a + b + c > 0) **then**

**begin**

**if** (x > 0) **then**

**begin**

**if** (y > x) **then**

**begin**

z := 5

**end**;

**end**;

**end**;

**end**.

Вывод:

**program** mfdsain;

**var** x:**integer**; y:**char**;

**begin**

**if** (a + b + c > 0) **then**

^ ^ ^

| | Переменная не была объявлена

| Переменная не была объявлена

Переменная не была объявлена

**begin**

**if** (x > 0) **then**

**begin**

**if** (y > x) **then**

^

Данную операцию нельзя применить к этим операндам

**begin**

z := 5

^

Переменная не была объявлена

**end**;

**end**;

**end**;

**end**.

6 Транслятор

6.1 Описание

Требуется написать генератор кода для программы Паскаль, генерировать код будем на языке CIL(MSIL). Common Intermediate Language (сокращённо CIL) — «высокоуровневый ассемблер» виртуальной машины .NET. Промежуточный язык, разработанный фирмой Microsoft для платформы .NET Framework. JIT-компилятор CIL является частью CLR (англ. common language runtime) — общей среды выполнения программ, написанных на языках .NET.

6.2 Проектирование

Создадим класс Translator, в конструктор которого будет передаваться синтаксически и семантически правильная в виде вектора токенов. Сразу после этого будет генерировать для нее код с помощью модуля System::Reflection::Assembly, для этого пришлось перенести проект на C++/CLI, чтобы воспользоваться модулями из C#.

Создадим приватную переменную – il\_generator. Она будет использоваться для создания тела методов и конструкторов в динамических сборках.

Главной функцией у il\_generator является функция Emit, с помощью нее можно положить инструкцию в поток инструкций, инструкции представлены в классе OpCodes.

В модуле транслятора создадим следующие методы:

* Добавление переменной
* Вызов инструкции исходя из оператора
* Вызов функции вывода в консоль(writeln)
* Вызов функции ввода из консоли(readln)
* Смену знака для значения на стеке
* Функция для загрузки значения константы в стек инструкций

Как видно, мы дополнительно ввели функции writeln и readln, в связи с этим пришлось дополнить синтаксический и семантический анализатор следующими БНФ:

* <readln>::= readln(<переменная>)
* <writeln>::= writeln(<выражение>)

Опкоды которые нам понадобятся:

|  |  |
| --- | --- |
| Название | Описание |
| Call | Вызывает метод, на который ссылается переданный дескриптор метода. |
| Stloc | Извлекает верхнее значение в стеке вычислений и сохраняет его в списке локальных переменных с заданным индексом. |
| Ret | Выполняет возврат из текущего метода, помещая возвращаемое значение (если имеется) из стека вычислений вызываемого метода в стек вычислений вызывающего метода. |
| Add | Складывает два значения и помещает результат в стек вычислений. |
| Sub | Вычитает одно значение из другого и помещает результат в стек вычислений. |
| Div | Делит одно значение на другое и помещает результат в стек вычислений как число с плавающей запятой (с типом F) или как частное (с типом int32). |
| Mul | Умножает два значения и помещает результат в стек вычислений. |
| Rem | Делит одно значение на другое и помещает остаток в стек вычислений. |
| And | Вычисляет побитовое И двух значений и помещает результат в стек вычислений. |
| Or | Вычисляет побитовое дополнение двух целочисленных значений, находящихся на вершине стека, и помещает результат в стек. |
| Not | Вычисляет побитовое дополнение целочисленного значения, находящегося на вершине стека, и помещает результат в стек с тем же типом. |
| Ceq | Сравнивает два значения. Если они равны, целочисленное значение 1 ((int32) помещается в стек вычислений; в противном случае в стек вычислений помещается 0 (int32). |
| Clt | Сравнивает два значения. Если первое значение меньше второго, целочисленное значение 1 ((int32) помещается в стек вычислений; в противном случае в стек вычислений помещается 0 (int32). |
| Cgt | Сравнивает два значения. Если первое значение больше второго, целочисленное значение 1 ((int32) помещается в стек вычислений; в противном случае в стек вычислений помещается 0 (int32). |
| Ldc\_I4 | Помещает переданное значение с типом int32 в стек вычислений как int32. |
| Ldc\_R8 | Помещает переданное значение с типом float64 в стек вычислений как F (число с плавающей запятой). |
| Ldstr | Помещает в стек ссылку на новый объект, представляющий строковой литерал, хранящийся в метаданных. |
| Ldloc | Загружает в стек вычислений локальную переменную с указанным индексом. |
| Brfalse | Передает управление конечной инструкции, если значением value является false, пустая ссылка (Nothing в Visual Basic) или ноль. |
| Br | Обеспечивает безусловную передачу управления конечной инструкции. |

6.3 Реализация

6.3.1 Функция добавления переменной

Dictionary<String^, LocalBuilder^> variable\_lb;

Dictionary<String^, System::Type^> variable\_type;

void Translator::add\_var(String^ name, Type^ type)

{

variable\_lb[name] = il\_generator->DeclareLocal(type);

variable\_type[name] = type;

}

6.3.2 Функция для вызова инструкции исходя из переданного оператора

void Translator::emit\_by\_operation(OperatorType op)

{

*/\**

*{"+", otPlus},*

*{"-", otMinus},*

*{"/", otSlash},*

*{"\*", otStar},*

*{"=", otEqual},*

*{"<", otLess},*

*{">", otGreater},*

*{">=", otGreaterEqual},*

*{"<=", otLessEqual},*

*{"<>", otLessGreater}\*/*

*/\**

*otDiv, // div*

*otAnd, // and*

*otNot, // not*

*otMod, // mod*

*otOr*

*\*/*

switch (op)

{

case otPlus: il\_generator->Emit(OpCodes::Add); break;

case otMinus: il\_generator->Emit(OpCodes::Sub); break;

case otSlash: il\_generator->Emit(OpCodes::Div); break;

case otStar: il\_generator->Emit(OpCodes::Mul); break;

case otDiv: il\_generator->Emit(OpCodes::Div); break;

case otMod: il\_generator->Emit(OpCodes::Rem); break;

case otAnd: il\_generator->Emit(OpCodes::And); break;

case otOr: il\_generator->Emit(OpCodes::Or); break;

case otNot: il\_generator->Emit(OpCodes::Not); break;

case otEqual: il\_generator->Emit(OpCodes::Ceq); break;

case otLess: il\_generator->Emit(OpCodes::Clt); break;

case otGreater: il\_generator->Emit(OpCodes::Cgt); break;

case otGreaterEqual:

il\_generator->Emit(OpCodes::Clt);

il\_generator->Emit(OpCodes::Ldc\_I4\_0);

il\_generator->Emit(OpCodes::Ceq);

break;

case otLessEqual:

il\_generator->Emit(OpCodes::Cgt);

il\_generator->Emit(OpCodes::Ldc\_I4\_0);

il\_generator->Emit(OpCodes::Ceq);

break;

case otLessGreater:

il\_generator->Emit(OpCodes::Ceq);

il\_generator->Emit(OpCodes::Ldc\_I4\_0);

il\_generator->Emit(OpCodes::Ceq);

break;

}

}

6.3.3 Функция вызова метода вывода в консоль(writeln)

void Translator::call\_writeln(System::Type^ type)

{

auto writeln\_method = Console::typeid->GetMethod("WriteLine", gcnew cli::array<Type^>{type});

il\_generator->Emit(OpCodes::Call, writeln\_method);

}

6.3.4 Функция вызова метода ввода из консоли(readln)

void Translator::call\_readln(Token\* token)

{

String^ name = get\_var\_name\_from\_token(token);

Type^ type = variable\_type[name];

LocalBuilder^ lb = variable\_lb[name];

MethodInfo^ readln\_method = Console::typeid->GetMethod("ReadLine", gcnew cli::array<Type^>{});

il\_generator->Emit(OpCodes::Call, readln\_method);

MethodInfo^ convert\_method = nullptr;

if (type == int::typeid)

{

convert\_method = Convert::typeid->GetMethod("ToInt32", gcnew cli::array<Type^>{string::typeid});

}

else if (type == double::typeid)

{

convert\_method = Convert::typeid->GetMethod("ToDouble", gcnew cli::array<Type^>{string::typeid});

}

else if (type == bool::typeid)

{

convert\_method = Convert::typeid->GetMethod("ToBoolean", gcnew cli::array<Type^>{string::typeid});

}

else if (type == char::typeid)

{

convert\_method = Convert::typeid->GetMethod("ToChar", gcnew cli::array<Type^>{string::typeid});

}

if (convert\_method != nullptr)

il\_generator->Emit(OpCodes::Call, convert\_method);

il\_generator->Emit(OpCodes::Stloc, lb);

}

6.3.5 Функция для смены знака значения на стеке

void Translator::change\_sign(System::Type^ type)

{

if (type == int::typeid) {

il\_generator->Emit(OpCodes::Ldc\_I4, -1);

il\_generator->Emit(OpCodes::Mul);

}

else if (type == double::typeid) {

il\_generator->Emit(OpCodes::Ldc\_R8, -1);

il\_generator->Emit(OpCodes::Mul);

}

}

6.3.6 Функция для загрузки константного значения в стек

void Translator::load\_const(Token\* token)

{

if (auto ct = dynamic\_cast<ConstToken<int>\*>(token)) {

il\_generator->Emit(OpCodes::Ldc\_I4, ct->value);

}

else if (auto ct = dynamic\_cast<ConstToken<double>\*>(token)) {

il\_generator->Emit(OpCodes::Ldc\_R8, ct->value);

}

else if (auto ct = dynamic\_cast<ConstToken<string>\*>(token)) {

il\_generator->Emit(OpCodes::Ldstr, convert\_to\_managed\_string(ct->value));

}

else if (auto ct = dynamic\_cast<ConstToken<char>\*>(token)) {

il\_generator->Emit(OpCodes::Ldstr, convert\_to\_managed\_string(string(1, ct->value)));

}

else if (auto ct = dynamic\_cast<ConstToken<bool>\*>(token)) {

il\_generator->Emit(OpCodes::Ldc\_I4, ct->value);

}

}

6.3.7 Полный код модуля транслятора

Заголовочный файл Translator.h:

#pragma once

#include "Token.h"

#include <string>

#include <vector>

#include <map>

#include <msclr\marshal\_cppstd.h>

using namespace std;

using namespace System;

using namespace System::Reflection::Emit;

using namespace System::Reflection;

using System::Reflection::AssemblyName;

using namespace System::Threading;

using namespace msclr::interop;

using namespace System::Collections::Generic;

ref class Translator

{

private:

int variable\_counter = 0;

ILGenerator^ il\_generator;

Dictionary<String^, LocalBuilder^> variable\_lb;

Dictionary<String^, System::Type^> variable\_type;

vector<Token\*>\* tokens = new vector<Token\*>();

int current\_token\_position;

Token\* current\_token;

OperatorType previous\_operator;

bool accept(TType token\_type);

bool accept(OperatorType operator\_type);

bool accept(vector<OperatorType> operator\_types);

void next\_token();

System::Type^ get\_type\_from\_token(Token\* token);

String^ get\_var\_name\_from\_token(Token\* token);

void add\_var(String^ name, System::Type^ type);

void emit\_by\_operation(OperatorType op);

String^ convert\_to\_managed\_string(string str);

void change\_sign(System::Type^ type);

void load\_const(Token\* token);

void call\_writeln(System::Type^ type);

void call\_readln(Token\* token);

void program();

void block();

void vars\_section();

void operators\_section();

bool single\_var\_definition();

System::Type^ type();

void operator\_();

bool simple\_operator();

System::Type^ expression();

bool relation\_operation();

System::Type^ simple\_expression();

bool additive\_operation();

System::Type^ term();

bool multiplicative\_operation();

System::Type^ factor();

bool sign();

void complex\_operator();

void neccessary\_compound\_operator();

bool compound\_operator();

bool if\_operator();

bool while\_operator();

bool writeln();

bool readln();

public:

Translator() {};

Translator(vector<Token\*> \_tokens);

~Translator();

};

Исходный файл Translator.cpp:

#include "pch.h"

#include "Translator.h"

Translator::Translator(vector<Token\*> \_tokens)

{

tokens = &\_tokens;

current\_token\_position = 0;

previous\_operator = otError;

next\_token();

String^ assemblyName = "MyProgram";

String^ modName = "MyProgram.exe";

String^ typeName = "MyProgram";

String^ methodName = "Main";

//Уникальный идентификатор сборки

AssemblyName^ name = gcnew AssemblyName(assemblyName);

//Домен приложения

AppDomain^ domain = Threading::Thread::GetDomain();

//Определяет и представляет динамическую сборку

AssemblyBuilder^ builder = domain->DefineDynamicAssembly(name, AssemblyBuilderAccess::RunAndSave);

//Определяет и представляет модуль в динамической сборке

ModuleBuilder^ module = builder->DefineDynamicModule(modName, true);

//Определяет и создает новые экземпляры классов во время выполнения

TypeBuilder^ typeBuilder = module->DefineType(typeName, TypeAttributes::Public);

//Определяет и представляет метод для динамического класса

MethodBuilder^ methodBuilder = typeBuilder->DefineMethod(methodName, MethodAttributes::Static

| MethodAttributes::Public, void::typeid, gcnew cli::array<Type^>{});

il\_generator = methodBuilder->GetILGenerator();

// generate

*/\**

*il\_generator->DeclareLocal(int::typeid);*

*il\_generator->Emit(OpCodes::Ldc\_I4, 10);*

*il\_generator->Emit(OpCodes::Stloc, 0);*

*auto x = il\_generator->DefineLabel();*

*auto y = il\_generator->DefineLabel();*

*il\_generator->MarkLabel(x);*

*il\_generator->Emit(OpCodes::Ldloc, 0);*

*il\_generator->Emit(OpCodes::Ldc\_I4, 100);*

*il\_generator->Emit(OpCodes::Clt);*

*il\_generator->Emit(OpCodes::Brfalse, y);*

*il\_generator->Emit(OpCodes::Ldloc, 0);*

*il\_generator->Emit(OpCodes::Ldc\_I4, 10);*

*il\_generator->Emit(OpCodes::Add);*

*il\_generator->Emit(OpCodes::Stloc, 0);*

*il\_generator->Emit(OpCodes::Ldloc, 0);*

*auto writeln\_test = Console::typeid->GetMethod("WriteLine", gcnew cli::array<Type^>{int::typeid});*

*il\_generator->Emit(OpCodes::Call, writeln\_test);*

*il\_generator->Emit(OpCodes::Br, x);*

*il\_generator->MarkLabel(y);*

*//typeof(Console).GetMethod("WriteLine", new Type[] { typeof(string) })*

*auto writeln\_test = Console::typeid->GetMethod("ReadLine", gcnew cli::array<Type^>{});*

*il\_generator->Emit(OpCodes::Call, writeln\_test);*

*auto lb = il\_generator->DeclareLocal(int::typeid);*

*il\_generator->Emit(OpCodes::Stloc, lb);*

*\*/*

program();

// end generate

// чтобы консоль не сразу закрывалась

il\_generator->EmitWriteLine(convert\_to\_managed\_string("Press any button..."));

auto readln\_method = Console::typeid->GetMethod("ReadLine", gcnew cli::array<Type^>{});

il\_generator->Emit(OpCodes::Call, readln\_method);

auto lb = il\_generator->DeclareLocal(String::typeid);

il\_generator->Emit(OpCodes::Stloc, lb);

//

il\_generator->Emit(OpCodes::Ret);

builder->SetEntryPoint(methodBuilder);

Type^ myClass = typeBuilder->CreateType();

//Сохраняем в \*.exe

builder->Save(modName);

}

Translator::~Translator()

{

delete il\_generator;

}

bool Translator::accept(TType token\_type)

{

bool result = true;

if (current\_token->token\_type != token\_type)

result = false;

if (result)

next\_token();

return result;

}

bool Translator::accept(OperatorType operator\_type)

{

bool result = true;

if (current\_token->token\_type != ttOperator)

result = false;

if (((OperatorToken\*)current\_token)->operator\_type != operator\_type)

result = false;

if (result)

next\_token();

return result;

}

bool Translator::accept(vector<OperatorType> operator\_types)

{

bool result = false;

if (current\_token->token\_type == ttOperator)

{

OperatorType current\_type = ((OperatorToken\*)current\_token)->operator\_type;

for (OperatorType operator\_type : operator\_types)

{

if (operator\_type == current\_type)

{

result = true;

break;

}

}

}

if (result)

next\_token();

return result;

}

void Translator::next\_token()

{

if (current\_token\_position == tokens->size())

return;

if (current\_token != NULL && current\_token->token\_type == ttOperator)

{

previous\_operator = ((OperatorToken\*)current\_token)->operator\_type;

}

current\_token = (\*tokens)[current\_token\_position];

current\_token\_position++;

}

String^ Translator::get\_var\_name\_from\_token(Token\* token)

{

string result = "";

if (token->token\_type == ttIdentificator)

result = ((IdentificatorToken\*)token)->name;

return convert\_to\_managed\_string(result);

}

Type^ Translator::get\_type\_from\_token(Token\* token)

{

if (auto ct = dynamic\_cast<ConstToken<int>\*>(token)) {

return int::typeid;

}

else if (auto ct = dynamic\_cast<ConstToken<double>\*>(token)) {

return double::typeid;

}

else if (auto ct = dynamic\_cast<ConstToken<string>\*>(token)) {

return string::typeid;

}

else if (auto ct = dynamic\_cast<ConstToken<char>\*>(token)) {

return char::typeid;

}

else

return bool::typeid;

}

void Translator::add\_var(String^ name, Type^ type)

{

variable\_lb[name] = il\_generator->DeclareLocal(type);

variable\_type[name] = type;

}

void Translator::emit\_by\_operation(OperatorType op)

{

*/\**

*{"+", otPlus},*

*{"-", otMinus},*

*{"/", otSlash},*

*{"\*", otStar},*

*{"=", otEqual},*

*{"<", otLess},*

*{">", otGreater},*

*{">=", otGreaterEqual},*

*{"<=", otLessEqual},*

*{"<>", otLessGreater}\*/*

*/\**

*otDiv, // div*

*otAnd, // and*

*otNot, // not*

*otMod, // mod*

*otOr*

*\*/*

switch (op)

{

case otPlus: il\_generator->Emit(OpCodes::Add); break;

case otMinus: il\_generator->Emit(OpCodes::Sub); break;

case otSlash: il\_generator->Emit(OpCodes::Div); break;

case otStar: il\_generator->Emit(OpCodes::Mul); break;

case otDiv: il\_generator->Emit(OpCodes::Div); break;

case otMod: il\_generator->Emit(OpCodes::Rem); break;

case otAnd: il\_generator->Emit(OpCodes::And); break;

case otOr: il\_generator->Emit(OpCodes::Or); break;

case otNot: il\_generator->Emit(OpCodes::Not); break;

case otEqual: il\_generator->Emit(OpCodes::Ceq); break;

case otLess: il\_generator->Emit(OpCodes::Clt); break;

case otGreater: il\_generator->Emit(OpCodes::Cgt); break;

case otGreaterEqual:

il\_generator->Emit(OpCodes::Clt);

il\_generator->Emit(OpCodes::Ldc\_I4\_0);

il\_generator->Emit(OpCodes::Ceq);

break;

case otLessEqual:

il\_generator->Emit(OpCodes::Cgt);

il\_generator->Emit(OpCodes::Ldc\_I4\_0);

il\_generator->Emit(OpCodes::Ceq);

break;

case otLessGreater:

il\_generator->Emit(OpCodes::Ceq);

il\_generator->Emit(OpCodes::Ldc\_I4\_0);

il\_generator->Emit(OpCodes::Ceq);

break;

}

}

void Translator::program() // <программа>::=program <имя>(<имя файла>{,<имя файла>});<блок>.

{

accept(otProgram);

accept(ttIdentificator);

accept(otSemiColon);

block();

accept(otDot);

}

String^ Translator::convert\_to\_managed\_string(string str)

{

return marshal\_as<String^>(str);

}

void Translator::block() // <блок>::=<раздел констант><раздел типов><раздел переменных><раздел процедур и функций><раздел операторов>

{

vars\_section();

operators\_section();

}

// ======== Раздел переменных ========

bool Translator::single\_var\_definition() // <описание однотипных переменных>::=<имя>{,<имя>}:<тип>

{

List<String^> variableNames;

variableNames.Add(get\_var\_name\_from\_token(current\_token));

if (!accept(ttIdentificator))

return false;

while (accept(otComma))

{

variableNames.Add(get\_var\_name\_from\_token(current\_token));

accept(ttIdentificator);

}

accept(otColon);

Type^ varType = type();

for (int i = 0; i < variableNames.Count; i++)

{

String^ name = variableNames[i];

add\_var(name, varType);

}

return true;

}

void Translator::call\_writeln(System::Type^ type)

{

auto writeln\_method = Console::typeid->GetMethod("WriteLine", gcnew cli::array<Type^>{type});

il\_generator->Emit(OpCodes::Call, writeln\_method);

}

void Translator::call\_readln(Token\* token)

{

String^ name = get\_var\_name\_from\_token(token);

Type^ type = variable\_type[name];

LocalBuilder^ lb = variable\_lb[name];

MethodInfo^ readln\_method = Console::typeid->GetMethod("ReadLine", gcnew cli::array<Type^>{});

il\_generator->Emit(OpCodes::Call, readln\_method);

MethodInfo^ convert\_method = nullptr;

if (type == int::typeid)

{

convert\_method = Convert::typeid->GetMethod("ToInt32", gcnew cli::array<Type^>{string::typeid});

}

else if (type == double::typeid)

{

convert\_method = Convert::typeid->GetMethod("ToDouble", gcnew cli::array<Type^>{string::typeid});

}

else if (type == bool::typeid)

{

convert\_method = Convert::typeid->GetMethod("ToBoolean", gcnew cli::array<Type^>{string::typeid});

}

else if (type == char::typeid)

{

convert\_method = Convert::typeid->GetMethod("ToChar", gcnew cli::array<Type^>{string::typeid});

}

if (convert\_method != nullptr)

il\_generator->Emit(OpCodes::Call, convert\_method);

il\_generator->Emit(OpCodes::Stloc, lb);

}

Type^ Translator::type() // <тип>::=integer|real|string|char

{

if (accept(otInteger))

return int::typeid;

if (accept(otReal))

return double::typeid;

if (accept(otString))

return String::typeid;

if (accept(otChar))

return Char::typeid;

}

void Translator::vars\_section() // <раздел переменных>::= var <описание однотипных переменных>;{<описание однотипных переменных>;} | <пусто>

{

accept(otVar);

single\_var\_definition();

accept(otSemiColon);

while (single\_var\_definition())

{

accept(otSemiColon);

}

}

// ======== Раздел операторов ========

// <раздел операторов>::= <составной оператор>

void Translator::operators\_section()

{

neccessary\_compound\_operator();

}

void Translator::change\_sign(System::Type^ type)

{

if (type == int::typeid) {

il\_generator->Emit(OpCodes::Ldc\_I4, -1);

il\_generator->Emit(OpCodes::Mul);

}

else if (type == double::typeid) {

il\_generator->Emit(OpCodes::Ldc\_R8, -1);

il\_generator->Emit(OpCodes::Mul);

}

}

void Translator::load\_const(Token\* token)

{

if (auto ct = dynamic\_cast<ConstToken<int>\*>(token)) {

il\_generator->Emit(OpCodes::Ldc\_I4, ct->value);

}

else if (auto ct = dynamic\_cast<ConstToken<double>\*>(token)) {

il\_generator->Emit(OpCodes::Ldc\_R8, ct->value);

}

else if (auto ct = dynamic\_cast<ConstToken<string>\*>(token)) {

il\_generator->Emit(OpCodes::Ldstr, convert\_to\_managed\_string(ct->value));

}

else if (auto ct = dynamic\_cast<ConstToken<char>\*>(token)) {

il\_generator->Emit(OpCodes::Ldstr, convert\_to\_managed\_string(string(1, ct->value)));

}

else if (auto ct = dynamic\_cast<ConstToken<bool>\*>(token)) {

il\_generator->Emit(OpCodes::Ldc\_I4, ct->value);

}

}

//<оператор>::=<простой оператор>|<сложный оператор>

void Translator::operator\_()

{

if (!simple\_operator())

complex\_operator();

}

//<простой оператор>::=<переменная>:=<выражение>

bool Translator::simple\_operator() // \*

{

int mem\_position = current\_token->position;

String^ name = get\_var\_name\_from\_token(current\_token);

if (!accept(ttIdentificator))

return false;

accept(otAssign);

Type^ t = expression();

// кладем значение

il\_generator->Emit(OpCodes::Stloc, variable\_lb[name]);

return true;

}

//<выражение>::=<простое выражение>|<простое выражение><операция отношения><простое выражение>

Type^ Translator::expression()

{

Type^ t1, ^ t2;

t1 = simple\_expression();

if (relation\_operation())

{

OperatorType mem\_op = previous\_operator;

t2 = simple\_expression();

emit\_by\_operation(mem\_op);

}

return t1;

}

//<операция отношения>::= =|<>|<|<=|>=|>

bool Translator::relation\_operation() // \*

{

return accept({ otEqual, otLessGreater, otLess, otLessEqual, otGreaterEqual, otGreater });

}

//<простое выражение>::=<слагаемое>{<аддитивная операция><слагаемое>}

Type^ Translator::simple\_expression()

{

Type ^t1, ^t2;

t1 = term();

while (additive\_operation())

{

OperatorType mem\_op = previous\_operator;

t2 = term();

emit\_by\_operation(mem\_op);

}

return t1;

}

//<аддитивная операция>::= +|-|or

bool Translator::additive\_operation() // \*

{

return accept({ otPlus, otMinus, otOr });

}

//<слагаемое>::=<множитель>{<мультипликативная операция><множитель>}

Type^ Translator::term()

{

Type^ t1, ^t2;

t1 = factor();

while (multiplicative\_operation())

{

OperatorType mem\_op = previous\_operator;

t2 = factor();

emit\_by\_operation(mem\_op);

}

return t1;

}

//<мультипликативная операция>::=\*|/|div|mod|and

bool Translator::multiplicative\_operation() // \*

{

return accept({ otStar, otSlash, otDiv, otMod, otAnd });

}

//<множитель>::=[<знак>]<переменная>|[<знак>]<константа>|[<знак>](<выражение>)|not <множитель>

Type^ Translator::factor()

{

// TODO: сделать что-то со знаком...

OperatorType mem\_op = otPlus;

if (sign())

{

mem\_op = previous\_operator;

}

Token\* mem\_token = current\_token;

if (accept(ttIdentificator))

{

String^ name = get\_var\_name\_from\_token(mem\_token);

il\_generator->Emit(OpCodes::Ldloc, variable\_lb[name]);

auto type = variable\_type[name];

if (mem\_op == otMinus)

change\_sign(type);

return type;

}

else if (accept(ttConst))

{

Type^ const\_type = get\_type\_from\_token(mem\_token);

load\_const(mem\_token);

if (mem\_op == otMinus)

change\_sign(const\_type);

return const\_type;

}

else if (accept(otLeftParenthesis))

{

Type^ type = expression();

accept(otRightParenthesis);

if (mem\_op == otMinus)

change\_sign(type);

return type;

}

else if (accept(otNot))

{

Type^ t = factor();

emit\_by\_operation(otNot);

return bool::typeid;

}

}

//<знак>::= +|-

bool Translator::sign() // \*

{

return accept({ otPlus, otMinus });

}

//<сложный оператор>::=<составной оператор>|<выбирающий оператор>|<оператор цикла>|<врайтлайн>

void Translator::complex\_operator()

{

if (compound\_operator())

{

//...

}

else if (if\_operator())

{

//...

}

else if (while\_operator())

{

//...

}

else if (writeln())

{

//...

}

else if (readln())

{

//...

}

}

//<составной оператор>::= begin <оператор>{;<оператор>} end

bool Translator::compound\_operator() // \*

{

if (!accept(otBegin))

return false;

operator\_();

while (accept(otSemiColon))

{

operator\_();

}

accept(otEnd);

return true;

}

//<обязательный составной оператор>::= begin <оператор>{;<оператор>} end

void Translator::neccessary\_compound\_operator() // \*

{

accept(otBegin);

operator\_();

while (accept(otSemiColon))

{

operator\_();

}

accept(otEnd);

}

//<выбирающий оператор>::= if <выражение> then <оператор>|if <выражение> then <оператор> else <оператор>

bool Translator::if\_operator() // \*

{

*/\**

*il\_generator->DeclareLocal(int::typeid);*

*il\_generator->Emit(OpCodes::Ldc\_I4, 10);*

*il\_generator->Emit(OpCodes::Stloc, 0);*

*auto x = il\_generator->DefineLabel();*

*auto y = il\_generator->DefineLabel();*

*il\_generator->MarkLabel(x);*

*il\_generator->Emit(OpCodes::Ldloc, 0);*

*il\_generator->Emit(OpCodes::Ldc\_I4, 100);*

*il\_generator->Emit(OpCodes::Clt);*

*il\_generator->Emit(OpCodes::Brfalse, y);*

*il\_generator->Emit(OpCodes::Ldloc, 0);*

*il\_generator->Emit(OpCodes::Ldc\_I4, 10);*

*il\_generator->Emit(OpCodes::Add);*

*il\_generator->Emit(OpCodes::Stloc, 0);*

*il\_generator->Emit(OpCodes::Ldloc, 0);*

*auto writeln\_test = Console::typeid->GetMethod("WriteLine", gcnew cli::array<Type^>{int::typeid});*

*il\_generator->Emit(OpCodes::Call, writeln\_test);*

*il\_generator->Emit(OpCodes::Br, x);*

*il\_generator->MarkLabel(y);\*/*

if (!accept(otIf))

return false;

auto then\_pointer = il\_generator->DefineLabel();

auto after\_pointer = il\_generator->DefineLabel();

Type^ t = expression();

il\_generator->Emit(OpCodes::Brfalse, then\_pointer);

accept(otThen);

operator\_();

il\_generator->Emit(OpCodes::Br, after\_pointer);

il\_generator->MarkLabel(then\_pointer);

if (accept(otElse))

{

operator\_();

}

il\_generator->MarkLabel(after\_pointer);

return true;

}

//<оператор цикла>::= while <выражение> do <оператор>

bool Translator::while\_operator() // \*

{

if (!accept(otWhile))

return false;

auto condition\_pointer = il\_generator->DefineLabel();

auto after\_pointer = il\_generator->DefineLabel();

il\_generator->MarkLabel(condition\_pointer);

Type^ t = expression();

il\_generator->Emit(OpCodes::Brfalse, after\_pointer);

accept(otDo);

operator\_();

il\_generator->Emit(OpCodes::Br, condition\_pointer);

il\_generator->MarkLabel(after\_pointer);

return true;

}

bool Translator::writeln()

{

if (!accept(otWriteLn))

return false;

accept(otLeftParenthesis);

Type^ t = expression();

accept(otRightParenthesis);

call\_writeln(t);

return true;

}

bool Translator::readln()

{

if (!accept(otReadLn))

return false;

accept(otLeftParenthesis);

Token\* mem\_token = current\_token;

accept(ttIdentificator);

call\_readln(mem\_token);

accept(otRightParenthesis);

return true;

}

6.4 Тестирование

Для просмотра инструкций сгенерированной программы воспользуемся инструментом дизассемблирования программы – ildasm.

6.4.1 Ввод и вывод числа

Исходный код программы:

**program** test;

**var** x:**integer**;

**begin**

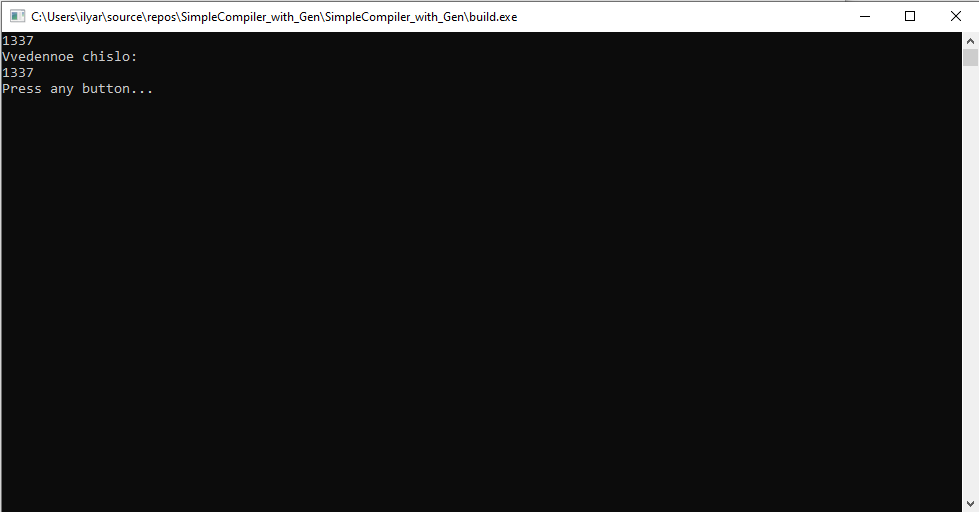
readln(x);

writeln("Vvedennoe chislo:");

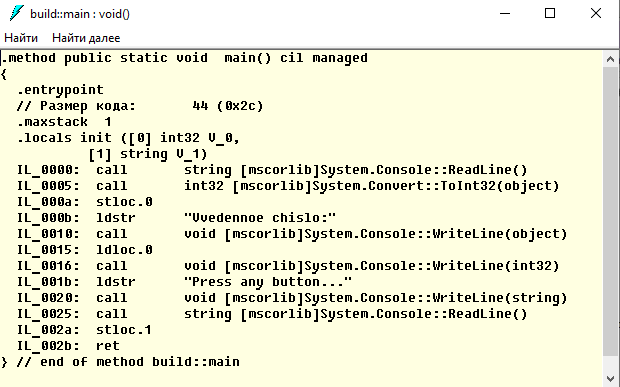
writeln(x)

**end**.

Работа программы:



Код на языке MSIL:



6.4.2 Проверка введенного числа на простоту

Исходный код программы:

**program** isprime;

**var** x,i,c:**integer**;

**begin**

writeln("Vvedite chislo");

readln(x);

c:=0;

i:=2;

**while** (i < x) **do**

**begin**

**if** (x **mod** i = 0) **then**

c := c + 1;

i := i+1

**end**;

**if** (c > 0) **then**

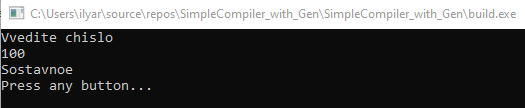
writeln("Sostavnoe")

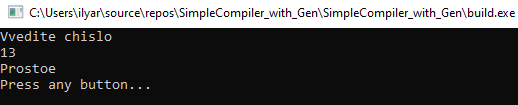
**else**

writeln("Prostoe")

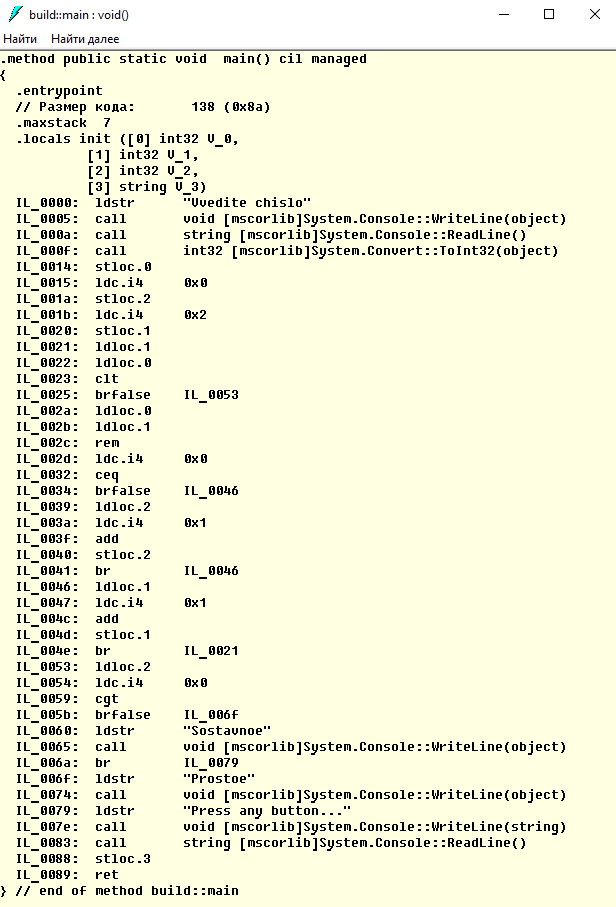
**end**.

Результат работы программы:





Код на языке MSIL:



6.4.2 Нахождение корня для полинома третьей степени

Исходный код:

**program** root\_p3;

**var** x,a,b,c,d,res,eps,l,r,m,fl,fr,fm:**real**;

**begin**

writeln("Vvedite koeficenti: a,b,c,d");

readln(a);

readln(b);

readln(c);

readln(d);

eps := 0.00000001;

l:=-1000.0;

r:=1000.0;

**while** (r - l > eps) **do**

**begin**

m := (l+r) / 2.0;

fm:= a \* m \* m \* m + b \* m \* m + c \* m + d;

fr:= a \* r \* r \* r + b \* r \* r + c \* r + d;

**if** (fr \* fm < 0.0) **then**

l := m

**else**

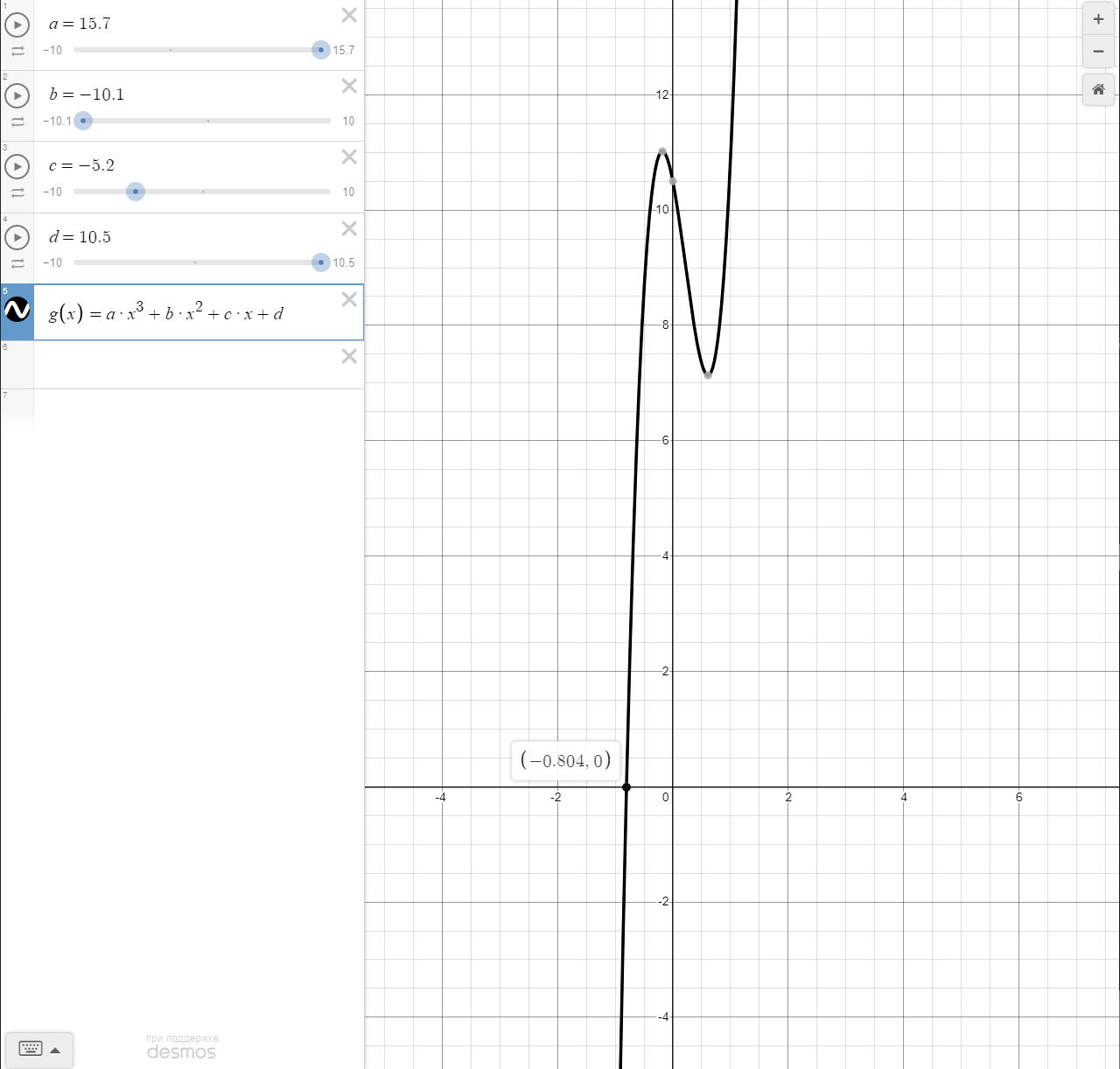
r := m

**end**;

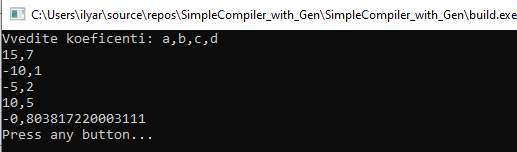
writeln((l+r)/2.0)

**end**.

Введем коэф-ы следующей функции:



Вывод программы:



Код программы на язык MSIL:

