Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«Владимирский государственный университет

имени Александра Григорьевича и Николая Григорьевича Столетовых»

(ВлГУ)

разработка компилятора подмножества

процедурно-ориентированного языка

Пояснительная записка

RU. 643.02068048.0001-01 81 01

# АННОТАЦИЯ

Компилятор – программный модуль, задачей которого является перевод программы, написанной на одном из языков программирования (исходный язык) в программу на машинном коде, либо в байт-код.

Курсовая работа заключается в создании компилятора подмножества процедурно-ориентированного языка. С исходным кодом проекта можно ознакомиться, перейдя по ссылке: <https://github.com/Elektro33rus/pascal-compiler>.

В пояснительной записке приведены требования к разрабатываемому проекту и описаны все стадии проектирования компилятора, а именно: построение лексического анализатора, создание синтаксического анализатора, проектирование генератора байт-кода, построение оптимизатора, а также приведены результаты тестирования разработанного компилятора.

**СОДЕРЖАНИЕ**

[АННОТАЦИЯ 2](#_Toc11595973)

[СОДЕРЖАНИЕ 3](#_Toc11595974)

[1 Требования к проекту 4](#_Toc11595975)

[2 ПРОЕКТИРОВАНИЕ ЛЕКСИЧЕСКОГО АНАЛИЗАТОРА 5](#_Toc11595976)

[2.1 Исходные данные 5](#_Toc11595977)

[2.2 Принципы работы лексического анализатора 5](#_Toc11595978)

[2.3 Пример работы лексического анализатора 8](#_Toc11595979)

[3 ПРОЕКТИРОВАНИЕ СИНТАКСИЧЕСКОГО АНАЛИЗАТОРА 9](#_Toc11595980)

[3.1 Исходные данные 9](#_Toc11595981)

[3.2 Принципы работы синтаксического анализатора 9](#_Toc11595982)

[3.3 Пример синтаксического разбора 14](#_Toc11595983)

[4 ПРОЕКТИРОВАНИЕ ГЕНЕРАТОРА БАЙТ КОДА 15](#_Toc11595984)

[4.1 Исходные данные 15](#_Toc11595985)

[4.2 Принцип работы генератора байт кода 15](#_Toc11595986)

[4.3 Пример работы генератора байт-кода 20](#_Toc11595987)

[5 РАЗРАБОТКА ОПТИМИЗАТОРА 21](#_Toc11595988)

[6 ТЕСТИРОВАНИЕ ГОТОВОГО ПРОЕКТА 22](#_Toc11595989)

[6.1 Пример 1 (все требования) 22](#_Toc11595990)

[6.2 Пример 2 (рекурсия) 27](#_Toc11595991)

[ЗАКЛЮЧЕНИЕ 30](#_Toc11595992)

[Приложение А 31](#_Toc11595993)

**1 Требования к проекту**

Курсовая работа заключается в создании компилятора с заданного подмножества языка Паскаль с незначительными модификациями и упрощениями. Результатами курсовой работы являются программная реализация заданного компилятора и пояснительная записка, оформленная в соответствии с требованиями ГОСТ.

Компилятор должен быть построен из следующих составных частей:

1. лексический анализатор;
2. синтаксический анализатор;
3. оптимизатор;
4. генератор машинного или байт кода.

Входной язык компилятора должен удовлетворять следующим требованиям:

* входная программа может быть разбита на строки произвольным образом, все пробелы и переводы строки должны игнорироваться компилятором;
* текст входной программы может содержать комментарии любой длины, которые должны игнорироваться компилятором;
* входная программа должна представлять собой единый модуль, но требуется предусмотреть вызов функций;
* должны быть предусмотрены следующие варианты операторов входной программы:

1) оператор присваивания (:=);

2) условный оператор (if)

3) оператор цикла (for, while)*;*

4) арифметические операции сложения (*+*), вычитания (*-*), деления (/), умножения (\*);

5) операции сравнения меньше (*<*), больше (*>*), равно (*=*), неравно (<>), больше или равно (>=), меньше или равно (<=);

6) логические операции «и» (*and*), «или» (*or*), «нет» (*not*);

7) break-continue;

8) Для изменения приоритета операций должны использоваться круглые скобки.

* два типа данных: целый (integer) и вещественный (real);
* базовый вывод переменных (writeln);

В качестве выходного (результирующего) языка должен использоваться язык ассемблера или байт-кода, был выбран последний.

# 2 ПРОЕКТИРОВАНИЕ ЛЕКСИЧЕСКОГО АНАЛИЗАТОРА

## **2.1 Исходные данные**

Для выполнения данной части курсовой работы требуется написать программу, которая выполняет лексический анализ входного текста в соответствии с заданием и порождает таблицу лексем с указанием их типов и значений. Программа должна выдавать сообщения о наличие во входном тексте ошибок, которые могут быть обнаружены на этапе лексического анализа.

Программа должна допускать наличие комментариев неограниченной длины во входном файле, которые игнорируются в дальнейшем.

## **2.2 Принципы работы лексического анализатора**

Запущенный лексический анализатор просматривает входной поток символов программы на исходном языке, выделяя символы, входящие в требуемую лексему, до обна­ружения очередного символа, который может ограничивать лексему.

В листинге 1 приведен код метода checkCharacter(char), с помощью которого производится разбор кода входной программы на лексемы. На данном этапе организовано игнорирование идентации текста и пробелов, выделение строк, чисел и операторов, а также реализована поддержка корректной обработки однострочных и многострочных комментариев. Все токены хранятся в структуре ArrayList<Token>.

Листинг 1 – Сканирование входного потока символов

**public** **static** **void** checkCharacter(**char** element){

**if** (*comment*) {

**if** (element == '}')

*comment* = **false**;

**if** (***CHAR\_TYPE***.get(String.*valueOf*(element)) == TYPE.***SPACE***){

*tokenName* = *endOfWord*();

**if** (element == Character.*toChars*(10)[0]){

*lineRow*++;

*lineCol* = 0;

} **else** **if** (element == Character.*toChars*(9)[0]){

*lineCol*+=4;

} **else** **if** (element == Character.*toChars*(32)[0]){

*lineCol*++;

}

}

**return**;

}

**switch** (***CHAR\_TYPE***.get(String.*valueOf*(element))){

**case** ***LETTER***:

**if** (!*readingNumber*) {

*tokenName* += element;

}

**if** (element == 'E' && *readingNumber*) {

*tokenName* += element;

*sciNotation* = **true**;

}

**break**;

**case** ***DIGIT***:

**if** (*tokenName*.isEmpty()) {

*readingNumber* = **true**;

}

*tokenName* += element;

**break**;

**case** ***SPACE***:

**if** (*readingString*){

*tokenName* += element;

} **else** **if** (*readingColon*) {

*generateToken*(***OPERATORS\_TOKEN***.get(*tokenName*));

*readingColon* = **false**;

} **else** **if** (*readingBool*) {

*generateToken*(***OPERATORS\_TOKEN***.get(*tokenName*));

*readingBool* = **false**;

} **else** **if** (!*readingNumber*) {

*tokenName* = *endOfWord*();

**if** (element == Character.*toChars*(10)[0]){

*lineRow*++;

*lineCol* = 0;

} **else** **if** (element == Character.*toChars*(9)[0]){

*lineCol*+=4;

} **else** **if** (element == Character.*toChars*(32)[0]){

*lineCol*++;

}

} **else** {

*handleNumber*();

}

**break**;

**case** ***OPERATOR***:

**if** (*readingDot* && element == '.') {

**if** (*tokenName*.equals(".")) {

*tokenName* = "";

*generateToken*("TK\_RANGE");

} **else** {

*generateToken*(*tokenName*.substring(0, *tokenName*.length()-2));

*generateToken*("TK\_DOT");

*tokenName* = "";

}

*readingDot* = **false**;

} **else** **if**(*readingString*) {

*tokenName* += element;

} **else** **if** (*readingNumber*) {

**if** (*isFloat* && element == '.') {

*isFloat* = **false**;

*tokenName* = *tokenName*.substring(0,*tokenName*.length()-1);

*handleNumber*();

*generateToken*("TK\_RANGE");

*tokenName* = "";

} **else** **if** (*sciNotation* && (element == '+' || element == '-')) {

*tokenName* += element;

} **else** **if** (element == '.') {

*isFloat* = **true**;

*tokenName* += element;

} **else** {

*handleNumber*();

*generateToken*(***OPERATORS\_TOKEN***.get(String.*valueOf*(element)));

}

} **else** **if** (*readingColon* && element == '=') {

*tokenName* += element;

*generateToken*(***OPERATORS\_TOKEN***.get(*tokenName*));

*readingColon* = **false**;

} **else** **if** (*readingBool*) {

**if** (*tokenName*.equals("<") && ((element == '=') || (element == '>'))) {

*tokenName* += element;

*generateToken*(***OPERATORS\_TOKEN***.get(*tokenName*));

} **else** **if** (*tokenName*.equals(">") && (element == '=')) {

*tokenName* += element;

*generateToken*(***OPERATORS\_TOKEN***.get(*tokenName*));

}

*readingBool* = **false**;

} **else** {

**if** (element == ';') {

*tokenName* = *endOfWord*();

*tokenName* = ";";

*generateToken*(***OPERATORS\_TOKEN***.get(String.*valueOf*(element)));

} **else** **if** (element == ':') {

*tokenName* = *endOfWord*();

*readingColon* = **true**;

*tokenName* += element;

} **else** **if** (element == '<' || element == '>') {

*tokenName* = *endOfWord*();

*readingBool* = **true**;

*tokenName* += element;

} **else** **if** (element == '.') {

*tokenName* += element;

**if** (*tokenName*.equals("end.")){

*generateToken*("TK\_END");

*generateToken*("TK\_DOT");

} **else** {

*readingDot* = **true**;

}

}

**else** **if** (***OPERATORS\_TOKEN***.containsKey(String.*valueOf*(element))) {

*tokenName* = *endOfWord*();

*tokenName* = String.*valueOf*(element);

*generateToken*(***OPERATORS\_TOKEN***.get(*tokenName*));

}

}

**break**;

**case** ***QUOTE***:

**if** (element == '{') {

*comment* = **true**;

**break**;

}

*readingString* = !*readingString*;

*tokenName* += element;

**if** (!*readingString*) {

*tokenName* = *tokenName*.substring(1, *tokenName*.length()-1);

**if** (*tokenName*.length() == 1) {

*generateToken*("TK\_CHARLIT");

} **else** **if** (*tokenName*.length() > 1) {

*generateToken*("TK\_STRLIT");

}

}

**break**;

**default**:

**throw** **new** Error("Unhandled element scanned");

}

}

Для распознавания входящих символов используются hash таблица, например для операторов (сложение, вычитание, сравнение, скобки и т.д.) используется hash таблица, указанная в листинге 2.

Листинг 2. Hash таблица операторов

**private** **static** **final** HashMap<String, String> ***OPERATORS\_TOKEN***;

**static** {

***OPERATORS\_TOKEN*** = **new** HashMap<>();

***OPERATORS\_TOKEN***.put("(", "TK\_OPEN\_PARENTHESIS");

***OPERATORS\_TOKEN***.put(")", "TK\_CLOSE\_PARENTHESIS");

***OPERATORS\_TOKEN***.put("[", "TK\_OPEN\_SQUARE\_BRACKET");

***OPERATORS\_TOKEN***.put("]", "TK\_CLOSE\_SQUARE\_BRACKET");

***OPERATORS\_TOKEN***.put(".", "TK\_DOT");

***OPERATORS\_TOKEN***.put("..", "TK\_RANGE");

***OPERATORS\_TOKEN***.put(":", "TK\_COLON");

***OPERATORS\_TOKEN***.put(";", "TK\_SEMI\_COLON");

***OPERATORS\_TOKEN***.put("+", "TK\_PLUS");

***OPERATORS\_TOKEN***.put("-", "TK\_MINUS");

***OPERATORS\_TOKEN***.put("\*", "TK\_MULTIPLY");

***OPERATORS\_TOKEN***.put("/", "TK\_DIVIDE");

***OPERATORS\_TOKEN***.put("<", "TK\_LESS\_THAN");

***OPERATORS\_TOKEN***.put("<=", "TK\_LESS\_THAN\_EQUAL");

***OPERATORS\_TOKEN***.put(">", "TK\_GREATER\_THAN");

***OPERATORS\_TOKEN***.put(">=", "TK\_GREATER\_THAN\_EQUAL");

***OPERATORS\_TOKEN***.put(":=", "TK\_ASSIGNMENT");

***OPERATORS\_TOKEN***.put(",", "TK\_COMMA");

***OPERATORS\_TOKEN***.put("=", "TK\_EQUAL");

***OPERATORS\_TOKEN***.put("<>", "TK\_NOT\_EQUAL");

}

Все зарезервированные слова (begin, end, while и т.д.) тоже находятся в hash таблице. Они берутся из файла keywords.txt.

## 

## **2.3 Пример работы лексического анализатора**

Рассмотрим ниже пример обработки текстового файла, код которого представлен в листинге 3.

Листинг 3. Пример входного файла

program example;

var y, n, x : integer;

begin

y:=0;

while (y < 10) do

begin

y:=y+1;

if (y = 5) then begin

continue;

end;

writeln(y);

end;

end.

На выходе получим набор токенов, которые показаны в листинге 4.

Листинг 4. Набор токенов

TK\_PROGRAM

TK\_IDENTIFIER

TK\_SEMI\_COLON

TK\_VAR

TK\_IDENTIFIER

TK\_COMMA

TK\_IDENTIFIER

TK\_COMMA

TK\_IDENTIFIER

TK\_COLON

TK\_INTEGER

TK\_SEMI\_COLON

TK\_BEGIN

TK\_IDENTIFIER

TK\_ASSIGNMENT

TK\_INTLIT

TK\_SEMI\_COLON

TK\_WHILE

TK\_OPEN\_PARENTHESIS

TK\_IDENTIFIER

TK\_LESS\_THAN

TK\_INTLIT

TK\_CLOSE\_PARENTHESIS

TK\_DO

TK\_BEGIN

TK\_IDENTIFIER

TK\_ASSIGNMENT

TK\_IDENTIFIER

TK\_PLUS

TK\_INTLIT

TK\_SEMI\_COLON

TK\_IF

TK\_OPEN\_PARENTHESIS

TK\_IDENTIFIER

TK\_EQUAL

TK\_INTLIT

TK\_CLOSE\_PARENTHESIS

TK\_THEN

TK\_BEGIN

TK\_IDENTIFIER

TK\_SEMI\_COLON

TK\_END

TK\_SEMI\_COLON

TK\_WRITELN

TK\_OPEN\_PARENTHESIS

TK\_IDENTIFIER

TK\_CLOSE\_PARENTHESIS

TK\_SEMI\_COLON

TK\_END

TK\_SEMI\_COLON

TK\_END

TK\_DOT

TK\_EOF

Полученные токены будут использоваться в следующем этапе парсера – синтаксическом

анализе.

# 3 ПРОЕКТИРОВАНИЕ СИНТАКСИЧЕСКОГО АНАЛИЗАТОРА

## **3.1 Исходные данные**

После получения массива токенов, следует распарсить их, создать представление, которое будет использоваться в следующем этапе компилятора. В качестве представления используется список команд, которые генератор байт-кода будет выполнять. КС-грамматика языка приведена в Приложении А.

## **3.2 Принципы работы синтаксического анализатора**

Синтаксический анализ файла с исходным кодом и составление списка команд производится в процессе просмотра каждого токена. В листинге 5 приведен код метода funcDeclaration(), который отвечает за разбор сигнатуры функции, а именно ее типа, имени и передаваемых в нее параметров, тела функции.

Листинг 5. Разбор сигнатуры функции

**private** **static** **void** funcDeclaration() {

**if** (*currentToken*.getTokenType().equals("TK\_FUNCTION")) {

*numberfunc*++;

*match*("TK\_FUNCTION");

*currentToken*.setTokenType("TK\_A\_FUNC");

String function = *currentToken*.getTokenValue();

String functionResult = *currentToken*.getTokenValue()+"result";

*match*("TK\_A\_FUNC");

*match*("TK\_OPEN\_PARENTHESIS");

*Region* = function;

**int** kolvoparametrov = *varDeclarationsFunc*();

*match*("TK\_CLOSE\_PARENTHESIS");

*match*("TK\_COLON");

String TK\_RESULT = *currentToken*.getTokenType();

**if** (TK\_RESULT.equals("TK\_INTEGER"))

*genOpCode*(OP\_CODE.***FUNCTIONSTARTINT***);

**else**

**if** (TK\_RESULT.equals("TK\_REAL"))

*genOpCode*(OP\_CODE.***FUNCTIONSTARTREAL***);

**else** {

String Row = String.*valueOf*(*currentToken*.getLineRow()+1);

String Col = String.*valueOf*(*currentToken*.getLineCol()+1);

**throw** **new** Error(String.*format*("Функция '"+function+"' возвращает неизвестный результат (может быть int или real)"+ "("+Row+" строка, "+Col+" столбец)"));

}

*getToken*();

*Region* = function;

Symbol symbolFunctionResult = **new** Symbol(functionResult,

"TK\_A\_RESULT\_VAR", *Region*,

***STRING\_TYPE\_HASH\_MAP***.get(TK\_RESULT.toLowerCase().substring(3)),

*dp*);

*dp* += 4;

**if** (SymbolTable.*lookup*(functionResult, symbolFunctionResult.getRegion()) == **null**) {

SymbolTable.*insert*(symbolFunctionResult);

*arraySymbols*.add(symbolFunctionResult);

*arraySymbols*.add(**new** Symbol("Stop", **null**, **null**, TYPE.***STOP***, 0));

}

*match*("TK\_SEMI\_COLON");

Symbol symbolFunction = **new** Symbol(function,

"TK\_A\_FUNC", "Global",

***STRING\_TYPE\_HASH\_MAP***.get(TK\_RESULT.toLowerCase().substring(3)),

*ip*);

symbolFunction.setNumber(*numberfunc*);

symbolFunction.setAmount(kolvoparametrov);

*Region* = function;

*declarations*();

**if** (SymbolTable.*lookup*(function, "Global") == **null**)

SymbolTable.*insert*(symbolFunction);

**else** {

String Row = String.*valueOf*(*currentToken*.getLineRow()+1);

String Col = String.*valueOf*(*currentToken*.getLineCol()+1);

**throw** **new** Error(String.*format*("Функция '"+function + "' уже объявлена!"+ "("+Row+" строка, "+Col+" столбец)"));

}

*match*("TK\_BEGIN");

*Region* = function;

*statements*();

*match*("TK\_END");

*match*("TK\_SEMI\_COLON");

**if** (TK\_RESULT.equals("TK\_INTEGER")) {

*genOpCode*(OP\_CODE.***FUNCTIONENDINT***);

*genAddress*(symbolFunctionResult.getAddress());

}

**else**

**if** (TK\_RESULT.equals("TK\_REAL"))

*genOpCode*(OP\_CODE.***FUNCTIONENDREAL***);

}

}

В соответствии с составленной КС-грамматикой необходимо произвести разбор тела функции (листинг 6). Когда текущий токен равен одному из операторов, парсер вызывает

функцию, которая описывает текущий оператор, это может цикл for, цикл while, условный if, присваивание, вывод числа, вызов функции.

Листинг 6. Разбор тела функций

**public** **static** **void** statements(){

**while**(!*currentToken*.getTokenType().equals("TK\_END")) {

**switch** (*currentToken*.getTokenType()) {

**case** "TK\_WHILE":

*whileStat*();

**break**;

**case** "TK\_IF":

*ifStat*();

**break**;

**case** "TK\_FOR":

*forStat*();

**break**;

**case** "TK\_WRITELN":

*writeStat*();

**break**;

**case** "TK\_A\_FUNC":

*callfunc*();

**break**;

**case** "TK\_IDENTIFIER":

Symbol symbol = *findSymbol*();

**if** (symbol != **null** && (symbol.getRegion().equals(*Region*) || symbol.getRegion().equals("Global") || symbol.getRegion().equals("FUNCVAR"+*Region*)))

*currentToken*.setTokenType(symbol.getTokenType());

**else** {

**if** (*currentToken*.getTokenValue().equals("break")) {

*genOpCode*(OP\_CODE.***BREAK***);

*match*("TK\_IDENTIFIER");

}

**else**

**if** (*currentToken*.getTokenValue().equals("continue")) {

*genOpCode*(OP\_CODE.***CONTINUE***);

*match*("TK\_IDENTIFIER");

}

**else** {

String Row = String.*valueOf*(*currentToken*.getLineRow()+1);

String Col = String.*valueOf*(*currentToken*.getLineCol()+1);

**throw** **new** Error(String.*format*("'"+*currentToken*.getTokenValue()+"' не объявлен ("+Row+" строка, "+Col+" столбец)"));

}

}

**break**;

**case** "TK\_A\_FUNC\_VAR":

*assignmentStat*();

**break**;

**case** "TK\_A\_VAR":

*assignmentStat*();

**break**;

**case** "TK\_A\_RESULT\_VAR":

*assignmentStat*();

**break**;

**case** "TK\_SEMI\_COLON":

*match*("TK\_SEMI\_COLON");

**break**;

**default**:

**return**;

}

}

}

Присваивание переменной какому-либо значению состоит из нескольких этапов. Первый – поиск символа текущей переменной в символьной таблице, если символ не найден, то выдается ошибка, что такое переменной не существует. Второй этап – проверка на соответствие типов, нельзя присвоить целочисленной переменной дробную часть, а чтобы присвоить вещественной переменной целочисленную переменную, вторую нужно преобразовать в вещественный тип. Функция для присваивания описана в листинге 7.

Листинг 7. Присваивание

**public** **static** **void** assignmentStat() {

Symbol symbol = *findSymbol*();

**if** (symbol != **null**) {

TYPE lhsType = symbol.getDataType();

**int** lhsAddress = symbol.getAddress();

**if** (*currentToken*.getTokenType().equals("TK\_A\_VAR"))

*match*("TK\_A\_VAR");

**else**

**if** (*currentToken*.getTokenType().equals("TK\_A\_RESULT\_VAR")) {

*match*("TK\_A\_RESULT\_VAR");

*genOpCode*(OP\_CODE.***FUNCRETURN***);

symbol.setResult(**true**);

}

**else**

**if** (*currentToken*.getTokenType().equals("TK\_A\_FUNC\_VAR"))

*match*("TK\_A\_FUNC\_VAR");

*match*("TK\_ASSIGNMENT");

TYPE rhsType = *E*();

**if** (lhsType == rhsType) {

*genOpCode*(OP\_CODE.***POP***);

*genAddress*(lhsAddress);

}

**else** **if** (lhsType == TYPE.***R*** && rhsType == TYPE.***I***) {

*genOpCode*(OP\_CODE.***CVR***);

*genOpCode*(OP\_CODE.***POP***);

*genAddress*(lhsAddress);

}

**else** {

String Row = String.*valueOf*(*currentToken*.getLineRow()+1);

String Col = String.*valueOf*(*currentToken*.getLineCol()+1);

**throw** **new** Error(String.*format*("Невозможно преобразовать тип (%s) в тип (%s ", lhsType, rhsType + ") ("+Row+" строка, "+Col+" столбец)"));

}

} **else** {

String Row = String.*valueOf*(*currentToken*.getLineRow()+1);

String Col = String.*valueOf*(*currentToken*.getLineCol()+1);

**throw** **new** Error(String.*format*("Неизвестная переменная '%s'", *currentToken*.getTokenValue()+ "("+Row+" строка, "+Col+" столбец)"));

}

}

Оператор ветвления должен содержать в себе цепочку обязательных токенов, а начинаться должен с ключевого слова “if”. Условный оператор if может быть как полноценным (с else), так и укороченный. Разбор оператора условного оператора ветвления if показан в листинге 8.

Листинг 8. Разбор условного оператора if

**public** **static** **void** ifStat(){

*match*("TK\_IF");

*match*("TK\_OPEN\_PARENTHESIS");

*genOpCode*(OP\_CODE.***IFCMP***);

*C*();

*match*("TK\_CLOSE\_PARENTHESIS");

*match*("TK\_THEN");

*match*("TK\_BEGIN");

*genOpCode*(OP\_CODE.***IFTHEN***);

*statements*();

*match*("TK\_END");

*match*("TK\_SEMI\_COLON");

**if**(*currentToken*.getTokenType().equals("TK\_ELSE")) {

*genOpCode*(OP\_CODE.***IFELSE***);

*match*("TK\_ELSE");

*match*("TK\_BEGIN");

*statements*();

*match*("TK\_END");

}

*genOpCode*(OP\_CODE.***IFEND***);

}

Так как в программе должен быть предусмотрен вызов функций, то каждая переменная должна иметь свой собственный регион объявления. Глобальные переменные – переменные, которые объявлены в самом начале, все остальные переменные будут иметь регион, в котором они инициализированы. Каждый раз, когда парсер находит новое объявление функции, регион меняется на имя этой функции. Если в функции будет вызвана переменная, которая имеет другой регион (кроме глобального) то парсер выдаст ошибку. Поиск текущей переменной представлен в листинге 9.

Листинг 9. Поиск переменной в областях видимости

**private** **static** Symbol findSymbol() {

Symbol symbol = SymbolTable.*lookup*(*currentToken*.getTokenValue(), "FUNCVAR"+*Region*);

**if** (symbol == **null**)

symbol = SymbolTable.*lookup*(*currentToken*.getTokenValue(), *Region*);

**if** (symbol == **null**)

symbol = SymbolTable.*lookup*(*currentToken*.getTokenValue(), "Global");

**return** symbol;

}

**private** **static** Symbol findSymbol(Token token) {

Symbol symbol = SymbolTable.*lookup*(token.getTokenValue(), "FUNCVAR"+*Region*);

**if** (symbol == **null**)

symbol = SymbolTable.*lookup*(token.getTokenValue(), *Region*);

**if** (symbol == **null**)

symbol = SymbolTable.*lookup*(token.getTokenValue(), "Global");

**return** symbol;

}

Сами переменные и функции находятся в символьной таблице (листинг 10).

Листинг 10. Символьная таблица

**public** **final** **class** SymbolTable {

**static** **class** Scope {

Symbol[] symbolTable = **new** Symbol[***HASH\_TABLE\_SIZE***];

Scope next = **null**;

}

**private** **static** **final** **int** ***HASH\_TABLE\_SIZE*** = 256;

**private** **static** Scope *headerScope* = **new** Scope();

**public** **static** **void** insert(Symbol symbol) {

**int** hashValueName = *hash*(symbol.getName());

**int** hashValueRegion = *hash*(symbol.getRegion());

**int** hash = hashValueName + hashValueRegion;

Symbol bucketCursor = *headerScope*.symbolTable[hash];

**if** (bucketCursor == **null**)

*headerScope*.symbolTable[hash] = symbol;

**else** {

**while** (bucketCursor.next != **null**)

bucketCursor = bucketCursor.next;

bucketCursor.next = symbol;

}

}

**public** **static** Symbol lookup(String symbolName, String regionName) {

**int** hashValueName = *hash*(symbolName);

**int** hashValueRegion = *hash*(regionName);

**int** hash = hashValueName + hashValueRegion;

Symbol bucketCursor = *headerScope*.symbolTable[hash];

Scope scopeCursor = *headerScope*;

**while** (scopeCursor != **null**) {

**while** (bucketCursor != **null**) {

**if** (bucketCursor.getName().equals(symbolName) && bucketCursor.getRegion().equals(regionName))

**return** bucketCursor;

bucketCursor = bucketCursor.next;

}

scopeCursor = scopeCursor.next;

}

**return** **null**;

}

**public** **static** **int** hash(String symbolName) {

**int** h = 0;

**for** (**int** i = 0; i < symbolName.length(); i++)

h = h + h + symbolName.charAt(i);

h\*=197;

**return** h = h % ***HASH\_TABLE\_SIZE***/2;

}

**public** **static** **void** openScope() {

Scope innerScope = **new** Scope();

innerScope.next = *headerScope*;

*headerScope* = innerScope;

}

**public** **static** **void** closeScope() {

*headerScope* = *headerScope*.next;

}

**public** **static** Scope getHeaderScope() {

**return** *headerScope*;

}

}

На выходе синтаксического анализатора получаем список команд, которые в дальнейшем используются при построении байт-кода. Список команд, которые генерирует парсер выглядит следующим образом:

***STARTPROGRAM***, ***FUNCTIONSTARTINT***, ***FUNCTIONSTARTREAL***, ***FUNCTIONENDINT***, ***FUNCTIONENDREAL***, ***STARTVARDECL***, ***INTVAR***, ***REALVAR***, ***COMMA***, ***ENDVARDECL***, ***PUSHVARFROMDECL***, ***HALT***, ***BREAK***, ***CONTINUE***, ***PUSHREAL***, ***PUSH***, ***PUSHFLOATLIT***, ***PUSHINTLIT***, ***PUSHINT***, ***FUNCTIONCALL***, ***FORSTART***, ***FORTO***, ***FORBEGIN***, ***FOREND***, ***WHILECMP***, ***WHILEBEGIN***, ***WHILEEND***, ***IFCMP***, ***IFTHEN***, ***IFELSE***, ***IFEND***, ***PRINT\_INT***, ***PRINT\_REAL***, ***PRINT\_NEWLINE***, ***FUNCRETURN***, ***POP***, ***AND***, ***OR***, ***PUSHVARFUNC***, ***ISCALLINT***, ***ISCALLREAL***, ***REPLACERESULT***, ***ADD***, ***XCHG***, ***CVR***, ***FADD***, ***SUB***, ***FSUB***, ***MULT***, ***FMULT***, ***FDIV***, ***DIV***, ***LSSIF***, ***LSS***, ***GTR***, ***LEQ***, ***GEQ***, ***EQL***, ***NEQL***

## **3.3 Пример синтаксического разбора**

Рассмотрим работу синтаксического анализатора при обработке файла, код которого представлен в листинге 11.

Листинг 11. Код входного файла

program example;

var y, n, x : integer;

begin

y:=0;

while (y < 10) do

begin

y:=y+1;

if (y = 5) then

begin

continue;

end;

writeln(y);

end;

end.

На выходе получим такой список команд:

OP\_CODE: STARTPROGRAM

OP\_CODE: PUSHINTLIT

OP\_CODE: POP

OP\_CODE: WHILECMP

OP\_CODE: PUSH

OP\_CODE: PUSHINTLIT

OP\_CODE: LSS

OP\_CODE: WHILEBEGIN

OP\_CODE: PUSH

OP\_CODE: PUSHINTLIT

OP\_CODE: ADD

OP\_CODE: POP

OP\_CODE: IFCMP

OP\_CODE: PUSH

OP\_CODE: PUSHINTLIT

OP\_CODE: EQL

OP\_CODE: IFTHEN

OP\_CODE: CONTINUE

OP\_CODE: IFEND

OP\_CODE: PUSH

OP\_CODE: PRINT\_INT

OP\_CODE: WHILEEND

OP\_CODE: HALT

**4 ПРОЕКТИРОВАНИЕ ГЕНЕРАТОРА БАЙТ КОДА**

## **4.1 Исходные данные**

Для выполнения заключительной части курсовой работы требуется написать программу, которая на основании списка команд синтаксического разбора порождает байт код. В качестве байт кода был выбран LLVM IR.

## **4.2 Принцип работы генератора байт кода**

Генератор байт-кода получает список команд от парсера и постепенно начинает составлять LLVM IR код. Главный метод генератора байт кода называется generate() (листинг 12), он смотрит, какую команду нужно выполнить и запускает функцию, которая записывает все операции этой команды в строку AllProgram. По достижению команды HALT, означающую конец команд, генератор прекращает свою работу. Для каждой инструкции есть свой метод, такие конструкции, как for, while, if состоят из нескольких команд, например оператор ветвления if состоит из таких команд, как IFCMP, IFTHEN, IFELSE, IFEND (Листинг 13).

Листинг 12. Метод generate(), отвечающий за выполнения команд

**public** **static** **void** generate() {

Parser.OP\_CODE opCode;

**do** {

opCode = *getOpCode*();

System.***out***.println(opCode);

**switch** (opCode) {

**case** ***PUSH***:

*push*();

**break**;

**case** ***ISCALLREAL***:

*callint* = **false**;

**break**;

**case** ***ISCALLINT***:

*callint* = **true**;

**break**;

**case** ***PUSHINT***:

*pushint*();

**break**;

**case** ***PUSHREAL***:

*pushreal*();

**break**;

**case** ***PUSHVARFUNC***:

*pushvarfunc*();

**break**;

**case** ***PUSHINTLIT***:

*pushi*();

**break**;

**case** ***PUSHFLOATLIT***:

*pushf*();

**break**;

**case** ***PUSHVARFROMDECL***:

*pushvarfromdecl*();

**break**;

**case** ***REPLACERESULT***:

*replaceresult*();

**break**;

**case** ***POP***:

*pop*();

**break**;

**case** ***CVR***:

*cvr*();

**break**;

**case** ***XCHG***:

*xchg*();

**break**;

**case** ***PRINT\_REAL***:

*printReal*();

**break**;

**case** ***PRINT\_INT***:

*printInt*();

**break**;

**case** ***PRINT\_NEWLINE***:

*printLn*();

**break**;

**case** ***HALT***:

*halt*();

**break**;

**case** ***EQL***:

*eqls*(*IsIf*, "eq");

**break**;

**case** ***NEQL***:

*eqls*(*IsIf*, "ne");

**break**;

**case** ***LSS***:

*eqls*(*IsIf*, "slt");

**break**;

**case** ***LEQ***:

*eqls*(*IsIf*, "sle");

**break**;

**case** ***GTR***:

*eqls*(*IsIf*, "sgt");

**break**;

**case** ***GEQ***:

*eqls*(*IsIf*, "sge");

**break**;

**case** ***WHILECMP***:

*whilecmp*();

**break**;

**case** ***WHILEBEGIN***:

String label6 = *stackNumber*.pop();

*AllProgram*+=";Label "+label6+"\n";

**break**;

**case** ***WHILEEND***:

*whileend*();

**break**;

**case** ***ADD***:

*add*();

**break**;

**case** ***FADD***:

*fadd*();

**break**;

**case** ***SUB***:

*sub*();

**break**;

**case** ***FSUB***:

*fsub*();

**break**;

**case** ***MULT***:

*mult*();

**break**;

**case** ***FMULT***:

*fmult*();

**break**;

**case** ***DIV***:

*div*();

**break**;

**case** ***FDIV***:

*fdiv*();

**break**;

**case** ***FUNCTIONSTARTINT***:

*funcstartint*();

*vardecl*="";

**break**;

**case** ***FUNCTIONSTARTREAL***:

*funcstartreal*();

**break**;

**case** ***FUNCTIONCALL***:

*funccall*();

**break**;

**case** ***FUNCRETURN***:

*funcreturn* = **true**;

**break**;

**case** ***FUNCTIONENDINT***:

*funcendint*();

**break**;

**case** ***FUNCTIONENDREAL***:

*funcendreal*();

**break**;

**case** ***STARTPROGRAM***:

*start*();

**break**;

**case** ***INTVAR***:

*vardecl*+="i32";

String varsDecl="%"+String.*valueOf*(*KolvoVar*);

*stackNumber*.push(varsDecl);

*KolvoVar*++;

**break**;

**case** ***COMMA***:

*vardecl*+=", ";

**break**;

**case** ***REALVAR***:

*KolvoVar*++;

*stackNumber*.push(String.*valueOf*(*KolvoVar*));

*vardecl*+="double";

**break**;

**case** ***STARTVARDECL***:

*KolvoVar*=0;

**break**;

**case** ***ENDVARDECL***:

*vardecl*+="){\n";

**break**;

**case** ***FORSTART***:

*forstart*();

**break**;

**case** ***FORTO***:

*forto*();

**break**;

**case** ***FORBEGIN***:

**break**;

**case** ***FOREND***:

*forend*();

**break**;

**case** ***IFCMP***:

*IsIf*=**true**;

**break**;

**case** ***IFTHEN***:

*ifthen*();

**break**;

**case** ***IFELSE***:

*ifelse*();

**break**;

**case** ***IFEND***:

*ifend*();

**break**;

**case** ***BREAK***:

*breaK*();

**break**;

**case** ***CONTINUE***:

*continuE*();

**break**;

**case** ***AND***:

String labelAnd = *stackNumber*.pop();

*AndOr*.push("AND");

*AllProgram*+=";Label "+labelAnd+" AND\n";

*stackNumber*.push(labelAnd);

**break**;

**case** ***OR***:

String labelOr = *stackNumber*.pop();

*AndOr*.push("OR");

*AllProgram*+=";Label "+labelOr+" OR\n";

*stackNumber*.push(labelOr);

**break**;

**case** ***STARTGLOBALVARS***:

*startglobalvars*();

**break**;

**case** ***STARTFUNCVARS***:

*startfuncvars*();

**break**;

**default**:

**throw** **new** Error(String.*format*("Unhandled case: %s", opCode));

}

}

**while** (opCode != Parser.OP\_CODE.***HALT***);

}

Листинг 13. Генератор LLVM IR для оператора ветвления if

**private** **static** **void** ifthen() {

*poryadok*.add("then");

**if** (!*dataArrayIf*.isEmpty()) {

*dataArrayVars* = *dataArrayIf*.pop();

*dataArrayIf*.push(*dataArrayVars*.clone());

}

**else**

*dataArrayIf*.push(*dataArrayVars*.clone());

String labelIfThen = *stackNumber*.pop();

*AllProgram*+=";Label "+labelIfThen+" ifthen\n";

}

**private** **static** **void** ifelse() {

*dataArrayVars* = *dataArrayIf*.pop();

*dataArrayIf*.push(*dataArrayVars*.clone());

*poryadok*.add("else");

String elseLabel = *giveMeVar*();

*ifStackLabels*.push(elseLabel);

String labelunknown = *getIfLabel*();

**if** (*IsNotBreak* && *IsNotContinue*) {

*AllProgram*+="br label "+labelunknown+"\n";

}

**else** {

**if** (!*IsNotBreak*)

*IsNotBreak* = **true**;

**if** (!*IsNotContinue*)

*IsNotContinue* = **true**;

}

*AllProgram*+="\n;Label "+elseLabel+" ifelse\n";

*ifStackLabels*.push(labelunknown);

*elseOn*.push(**true**);

}

**private** **static** **void** ifend() {

String labelexit = *giveMeVar*();

String criptoExitStart=*ifStackLabels*.pop();

**if** (*ifStackLabels*.isEmpty()) {

*AllProgram*=*AllProgram*.replace(criptoExitStart, labelexit);

}

**else** {

**if** (!*elseOn*.isEmpty() && *ifStackLabels*.size()==3) {

String labelElse = *ifStackLabels*.pop();

String criptoExitThen = *ifStackLabels*.pop();

String criptoExitElse = *ifStackLabels*.pop();

*AllProgram*=*AllProgram*.replace(criptoExitStart, labelexit);

*AllProgram*=*AllProgram*.replace(criptoExitThen, labelElse);

*AllProgram*=*AllProgram*.replace(criptoExitElse, labelexit);

*elseOn*.pop();

} **else** {

String labelElse = *ifStackLabels*.pop();

String criptoExitThen = *ifStackLabels*.pop();

*AllProgram*=*AllProgram*.replace(criptoExitStart, labelexit);

*AllProgram*=*AllProgram*.replace(criptoExitThen, labelElse);

}

}

**if** (*IsNotBreak* && *IsNotContinue*) {

*AllProgram*+="br label "+labelexit+"\n";

}

**else** {

**if** (!*IsNotBreak*)

*IsNotBreak* = **true**;

**if** (!*IsNotContinue*)

*IsNotContinue* = **true**;

}

*AllProgram*+="\n;Label "+labelexit+" ifend\n";

*dataArrayVars* = *dataArrayIf*.pop();

}

После завершения генерации байт-кода он записывается в файл ouput с расширением ll. После этого его можно запустить через команду lli, либо преобразовать в объектный файл, используя команду llc, а затем слинковать для получения исполняемого файла.

## **4.3 Пример работы генератора байт-кода**

Рассмотрим работу генератора байт-кода при обработке файла, код которого приведен в листинге 14.

Листинг 14. Пример входного файла

program example;

var y, n, x : integer;

begin

y:=0;

while (y < 10) do

begin

y:=y+1;

if (y = 5) then

begin

continue;

end;

writeln(y);

end;

end.

После завершения, генератор выдает код, который показан в листинге 15.

Листинг 15. Полученный LLVM IR код

@.strln = private unnamed\_addr constant [2 x i8] c"\0A\00"

@.strint = private unnamed\_addr constant [4 x i8] c"%i\0A\00"

@.strfloat = private unnamed\_addr constant [4 x i8] c"%f\0A\00"

declare i32 @printf(i8\*, ...)

define i32 @main() {

%1 = alloca i32

store i32 0, i32\* %1

%2 = load i32, i32\* %1

store i32 %2, i32\* %1

br label %3

;Label %3

%4 = alloca i32

store i32 10, i32\* %4

%5 = load i32, i32\* %1

%6 = load i32, i32\* %4

%7 = icmp slt i32 %5, %6

br i1 %7, label %8, label %23

;Label %8

%9 = alloca i32

store i32 1, i32\* %9

%10 = load i32, i32\* %9

%11 = load i32, i32\* %1

%12 = add i32 %10, %11

%13 = alloca i32

store i32 %12, i32\* %13

%14 = load i32, i32\* %13

store i32 %14, i32\* %1

%15 = alloca i32

store i32 5, i32\* %15

%16 = load i32, i32\* %1

%17 = load i32, i32\* %15

%18 = icmp eq i32 %16, %17

br i1 %18, label %19, label %20

;Label %19 ifthen

br label %3

;Label %20 ifend

%21 = load i32, i32\* %1

%22 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strint, i32 0, i32 0), i32 %21)

br label %3

;Label %23

ret i32 0

}

Запуск программы с помощью команды lli показан на рисунке 1.

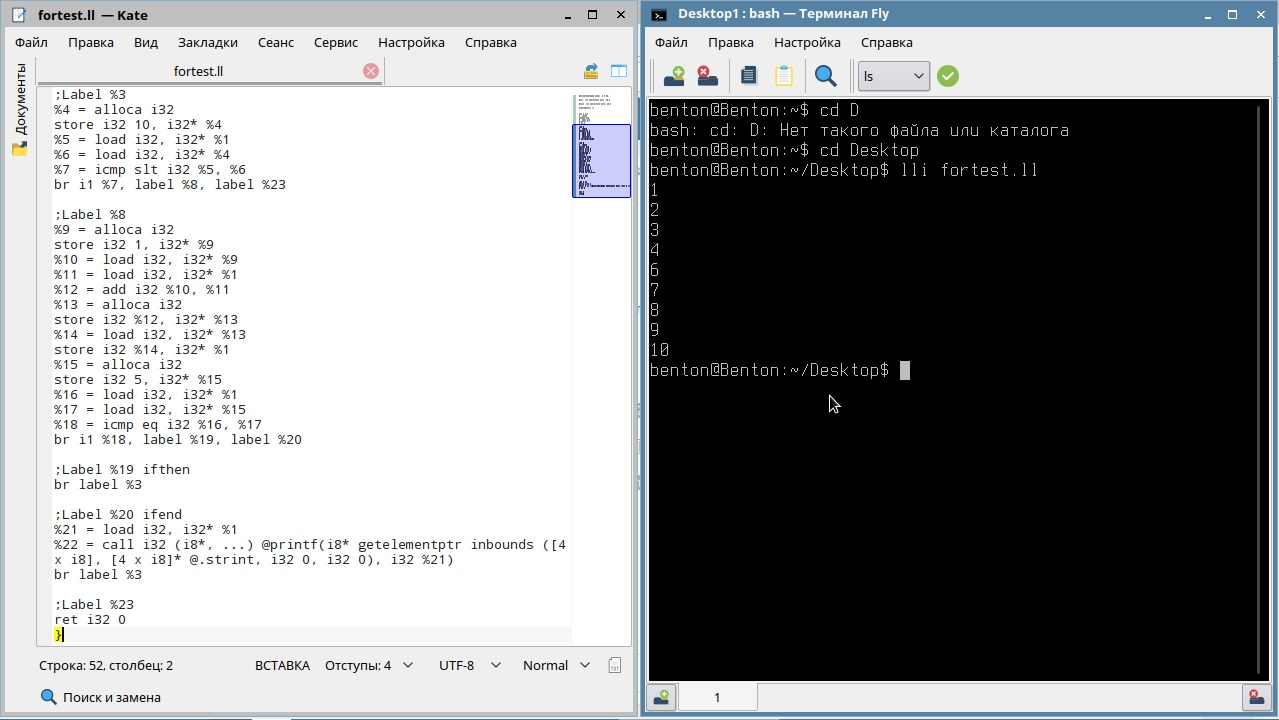


Рисунок 1. Запуск программы

Как видим, полученный код на LLVM IR компилируется без ошибок и выводит верный результат.

# 5 РАЗРАБОТКА ОПТИМИЗАТОРА

Оптимизация кода начинается после генерации байт-кода. Оптимизатор удаляет из кода функции, которые ни разу не вызывались, удаляет переменные, которые не использовались.

Проверка на использование переменной осуществляется за счёт флага, который выставляется во время генерации байт-кода.

# 6 ТЕСТИРОВАНИЕ ГОТОВОГО ПРОЕКТА

## **6.1 Пример 1 (все требования)**

Пример, который описывает все необходимые требования курсовой работы, приведен в листинге 16. Он сочетает в себе всё, что требовалось реализовать в задании курсовой работы.

Листинг 16.Пример со всеми требованиями

program example;

var x, i, k, d, e: integer; {два типа данных целочисленный}

var z, l: real; {два типа данных вещественный}

{комментарий

любой длины который

игнорируется компилятором!!!}

function test(): integer; {функция}

var y: integer;

begin

testresult:=100; {оператор присваивания}

y:=

10; {идентации не учитываются}

writeln(y); {базовый вывод переменных}

end;

function test2(round : integer): integer; {функция с переменной}

begin

test2result:=0;

writeln(round);

end;

begin

x:=test(); {вызов функции}

writeln(x); {функция возвращает результат}

test2(5); {вызов функции с переменной}

e:=1;

if (5 < x) then {условный оператор с else}

begin

writeln(1);

end;

else begin

writeln(0);

end;

for i:=0 to 10 do {оператор цикла for}

begin

if (i = 5) then {условный оператор без else}

begin

writeln(666);

break; {оператор break}

end;

writeln(i);

end;

k:=(50+100)\*2-(30\*1+x)\*2; {изменение приоритета операций}

writeln(k);

d:=20;

while (d > 10 ) do {оператор цикла while}

begin

writeln(d);

d:=d-1;

end;

while (e <= 16 ) do {оператор continue}

begin

e:=e\*2;

if (e = 4) then

begin continue;

end;

else

begin

writeln(e);

end;

end; z:=50.16+e\*2+100/5; {переменная real}

writeln(z);

end.

Компилятор переводит код программы из листинга 16 в байт код LLVM IR (листинг 17).

Листинг 17. Полученный LLVM IR

@.strln = private unnamed\_addr constant [2 x i8] c"\0A\00"

@.strint = private unnamed\_addr constant [4 x i8] c"%i\0A\00"

@.strfloat = private unnamed\_addr constant [4 x i8] c"%f\0A\00"

declare i32 @printf(i8\*, ...)

@"x" = internal global i32 undef

@"i" = internal global i32 undef

@"k" = internal global i32 undef

@"d" = internal global i32 undef

@"e" = internal global i32 undef

@"z" = internal global double undef

@"l" = internal global double undef

define i32 @Func0(){

%"testresult" = alloca i32

%"y" = alloca i32

%1 = alloca i32

store i32 100, i32\* %1

%2 = load i32, i32\* %1

store i32 %2, i32\* %"testresult"

%3 = alloca i32

store i32 10, i32\* %3

%4 = load i32, i32\* %3

store i32 %4, i32\* %"y"

%5 = load i32, i32\* %"y"

%6 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strint, i32 0, i32 0), i32 %5)

%7 = load i32, i32\* %"testresult"

ret i32 %7

}

define i32 @Func1(i32){

%2 = alloca i32

store i32 %0, i32\* %2

%"test2result" = alloca i32

%3 = alloca i32

store i32 0, i32\* %3

%4 = load i32, i32\* %3

store i32 %4, i32\* %"test2result"

%5 = load i32, i32\* %2

%6 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strint, i32 0, i32 0), i32 %5)

%7 = load i32, i32\* %"test2result"

ret i32 %7

}

define i32 @main() {

%1 = call i32 @Func0()

%2 = alloca i32

store i32 %1, i32\* %2

%3 = load i32, i32\* %2

store i32 %3, i32\* @"x"

%4 = load i32, i32\* @"x"

%5 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strint, i32 0, i32 0), i32 %4)

%6 = alloca i32

store i32 5, i32\* %6

%7 = load i32, i32\* %6

%8 = call i32 @Func1(i32 %7)

%9 = alloca i32

store i32 %8, i32\* %9

%10 = alloca i32

store i32 5, i32\* %10

%11 = load i32, i32\* %10

%12 = load i32, i32\* @"x"

%13 = icmp slt i32 %11, %12

br i1 %13, label %14, label %18

;Label %14 ifthen

%15 = alloca i32

store i32 1, i32\* %15

%16 = load i32, i32\* %15

%17 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strint, i32 0, i32 0), i32 %16)

br label %22

;Label %18 ifelse

%19 = alloca i32

store i32 0, i32\* %19

%20 = load i32, i32\* %19

%21 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strint, i32 0, i32 0), i32 %20)

br label %22

;Label %22 ifend

%23 = alloca i32

store i32 0, i32\* %23

%24 = load i32, i32\* %23

store i32 %24, i32\* @"i"

;ForStart

%25 = load i32, i32\* @"i"

%26 = alloca i32

store i32 %25, i32\* %26

br label %27

;label (forcmp) %27

%28 = alloca i32

store i32 10, i32\* %28

%29 = load i32, i32\* %26

%30 = load i32, i32\* %28

%31 = icmp sle i32 %29, %30

br i1 %31, label %32, label %47

;Label (fordo) %32

%33 = alloca i32

store i32 5, i32\* %33

%34 = load i32, i32\* %26

%35 = load i32, i32\* %33

%36 = icmp eq i32 %34, %35

br i1 %36, label %37, label %41

;Label %37 ifthen

%38 = alloca i32

store i32 666, i32\* %38

%39 = load i32, i32\* %38

%40 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strint, i32 0, i32 0), i32 %39)

br label %47

;Label %41 ifend

%42 = load i32, i32\* %26

%43 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strint, i32 0, i32 0), i32 %42)

br label %44

;Label (+1) %44

%45 = load i32, i32\* %26

%46 = add nsw i32 %45, 1

store i32 %46, i32\* %26

br label %27

;label (forend) %47

%48 = alloca i32

store i32 50, i32\* %48

%49 = alloca i32

store i32 100, i32\* %49

%50 = load i32, i32\* %49

%51 = load i32, i32\* %48

%52 = add i32 %50, %51

%53 = alloca i32

store i32 %52, i32\* %53

%54 = alloca i32

store i32 2, i32\* %54

%55 = load i32, i32\* %53

%56 = load i32, i32\* %54

%57 = mul i32 %55, %56

%58 = alloca i32

store i32 %57, i32\* %58

%59 = alloca i32

store i32 30, i32\* %59

%60 = alloca i32

store i32 1, i32\* %60

%61 = load i32, i32\* %59

%62 = load i32, i32\* %60

%63 = mul i32 %61, %62

%64 = alloca i32

store i32 %63, i32\* %64

%65 = load i32, i32\* @"x"

%66 = load i32, i32\* %64

%67 = add i32 %65, %66

%68 = alloca i32

store i32 %67, i32\* %68

%69 = alloca i32

store i32 2, i32\* %69

%70 = load i32, i32\* %68

%71 = load i32, i32\* %69

%72 = mul i32 %70, %71

%73 = alloca i32

store i32 %72, i32\* %73

%74 = load i32, i32\* %58

%75 = load i32, i32\* %73

%76 = sub i32 %74, %75

%77 = alloca i32

store i32 %76, i32\* %77

%78 = load i32, i32\* %77

store i32 %78, i32\* @"k"

%79 = load i32, i32\* @"k"

%80 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strint, i32 0, i32 0), i32 %79)

%81 = alloca i32

store i32 20, i32\* %81

%82 = load i32, i32\* %81

store i32 %82, i32\* @"d"

br label %83

;Label %83

%84 = alloca i32

store i32 10, i32\* %84

%85 = load i32, i32\* @"d"

%86 = load i32, i32\* %84

%87 = icmp sgt i32 %85, %86

br i1 %87, label %88, label %97

;Label %88

%89 = load i32, i32\* @"d"

%90 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strint, i32 0, i32 0), i32 %89)

%91 = alloca i32

store i32 1, i32\* %91

%92 = load i32, i32\* @"d"

%93 = load i32, i32\* %91

%94 = sub i32 %92, %93

%95 = alloca i32

store i32 %94, i32\* %95

%96 = load i32, i32\* %95

store i32 %96, i32\* @"d"

br label %83

;Label %97

%98 = alloca i32

store i32 1, i32\* %98

%99 = load i32, i32\* %98

store i32 %99, i32\* @"e"

br label %100

;Label %100

%101 = alloca i32

store i32 16, i32\* %101

%102 = load i32, i32\* @"e"

%103 = load i32, i32\* %101

%104 = icmp sle i32 %102, %103

br i1 %104, label %105, label %121

;Label %105

%106 = alloca i32

store i32 2, i32\* %106

%107 = load i32, i32\* @"e"

%108 = load i32, i32\* %106

%109 = mul i32 %107, %108

%110 = alloca i32

store i32 %109, i32\* %110

%111 = load i32, i32\* %110

store i32 %111, i32\* @"e"

%112 = alloca i32

store i32 4, i32\* %112

%113 = load i32, i32\* @"e"

%114 = load i32, i32\* %112

%115 = icmp eq i32 %113, %114

br i1 %115, label %116, label %117

;Label %116 ifthen

br label %100

;Label %117 ifelse

%118 = load i32, i32\* @"e"

%119 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strint, i32 0, i32 0), i32 %118)

br label %120

;Label %120 ifend

br label %100

;Label %121

%122 = alloca double

store double 50.16, double\* %122

%123 = alloca i32

store i32 2, i32\* %123

%124 = load i32, i32\* @"e"

%125 = load i32, i32\* %123

%126 = mul i32 %124, %125

%127 = alloca i32

store i32 %126, i32\* %127

%128 = load i32, i32\* %127

%129 = sitofp i32 %128 to double

%130 = alloca double

store double %129, double\* %130

%131 = load double, double\* %130

%132 = load double, double\* %122

%133 = fadd double %131, %132

%134 = alloca double

store double %133, double\* %134

%135 = alloca i32

store i32 100, i32\* %135

%136 = alloca i32

store i32 5, i32\* %136

%137 = load i32, i32\* %136

%138 = sitofp i32 %137 to double

%139 = alloca double

store double %138, double\* %139

%140 = load i32, i32\* %135

%141 = sitofp i32 %140 to double

%142 = alloca double

store double %141, double\* %142

%143 = load double, double\* %142

%144 = load double, double\* %139

%145 = fdiv double %143, %144

%146 = alloca double

store double %145, double\* %146

%147 = load double, double\* %146

%148 = load double, double\* %134

%149 = fadd double %147, %148

%150 = alloca double

store double %149, double\* %150

%151 = load double, double\* %150

store double %151, double\* @"z"

%152 = load double, double\* @"z"

%153 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strfloat, i32 0, i32 0), double %152)

ret i32 0

}

Теперь полученный LLVM IR можно запустить с помощью команды lli (рисунок 2).

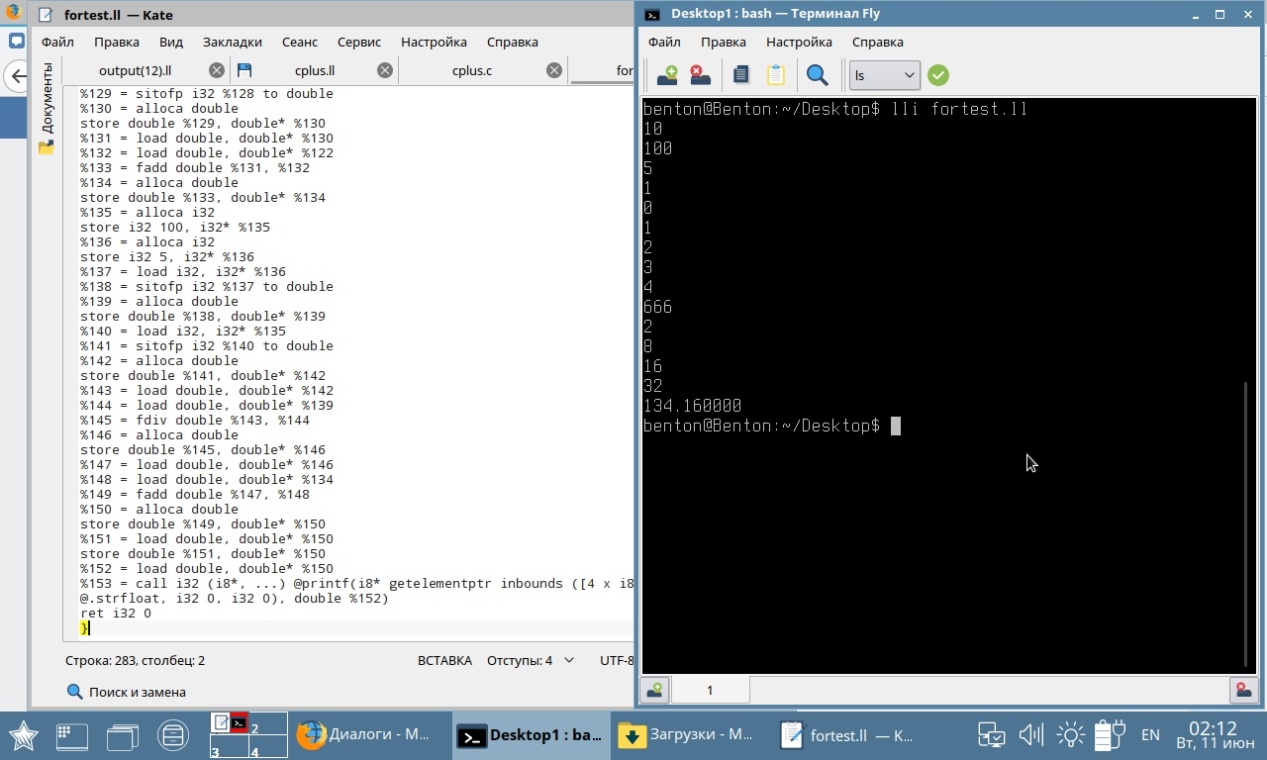


Рисунок 2. Запуск LLVM IR кода

## **6.2 Пример 2 (рекурсия)**

Компилятор умеет выполнять рекурсию, для примера возьмём рекурсию для подсчёта факториала (листинг 18). Она считает факториал числа 10.

Листинг 18. Код для подсчёта факториала

program example;

var x, n: integer;

function fact(a : integer): integer;

var temp1, temp2: integer;

begin

if (a <= 1) then

begin

a:=1;

end;

else

begin

temp1:=a-1;

temp2:=fact(temp1);

a:=a\*temp2;

end;

factresult:=a;

end;

begin

n:=10;

x:=fact(n);

writeln(x);

end.

Полученный код LLVM IR продемонстрирован в листинге 19.

Листинг 19. LLVM IR

@.strln = private unnamed\_addr constant [2 x i8] c"\0A\00"

@.strint = private unnamed\_addr constant [4 x i8] c"%i\0A\00"

@.strfloat = private unnamed\_addr constant [4 x i8] c"%f\0A\00"

declare i32 @printf(i8\*, ...)

@"x" = internal global i32 undef

@"n" = internal global i32 undef

define i32 @Func0(i32){

%2 = alloca i32

store i32 %0, i32\* %2

%"factresult" = alloca i32

%"temp1" = alloca i32

%"temp2" = alloca i32

%3 = alloca i32

store i32 1, i32\* %3

%4 = load i32, i32\* %2

%5 = load i32, i32\* %3

%6 = icmp sle i32 %4, %5

br i1 %6, label %7, label %10

;Label %7 ifthen

%8 = alloca i32

store i32 1, i32\* %8

%9 = load i32, i32\* %8

store i32 %9, i32\* %2

br label %26

;Label %10 ifelse

%11 = alloca i32

store i32 1, i32\* %11

%12 = load i32, i32\* %2

%13 = load i32, i32\* %11

%14 = sub i32 %12, %13

%15 = alloca i32

store i32 %14, i32\* %15

%16 = load i32, i32\* %15

store i32 %16, i32\* %"temp1"

%17 = load i32, i32\* %"temp1"

%18 = call i32 @Func0(i32 %17)

%19 = alloca i32

store i32 %18, i32\* %19

%20 = load i32, i32\* %19

store i32 %20, i32\* %"temp2"

%21 = load i32, i32\* %2

%22 = load i32, i32\* %"temp2"

%23 = mul i32 %21, %22

%24 = alloca i32

store i32 %23, i32\* %24

%25 = load i32, i32\* %24

store i32 %25, i32\* %2

br label %26

;Label %26 ifend

%27 = load i32, i32\* %2

store i32 %27, i32\* %"factresult"

%28 = load i32, i32\* %"factresult"

ret i32 %28

}

define i32 @main() {

%1 = alloca i32

store i32 10, i32\* %1

%2 = load i32, i32\* %1

store i32 %2, i32\* @"n"

%3 = load i32, i32\* @"n"

%4 = call i32 @Func0(i32 %3)

%5 = alloca i32

store i32 %4, i32\* %5

%6 = load i32, i32\* %5

store i32 %6, i32\* @"x"

%7 = load i32, i32\* @"x"

%8 = call i32 (i8\*, ...) @printf(i8\* getelementptr inbounds ([4 x i8], [4 x i8]\* @.strint, i32 0, i32 0), i32 %7)

ret i32 0

}

Если запустить его командой lli, в консоле появится число 3628800 (рисунок 3), это действительно факториал числа 10.

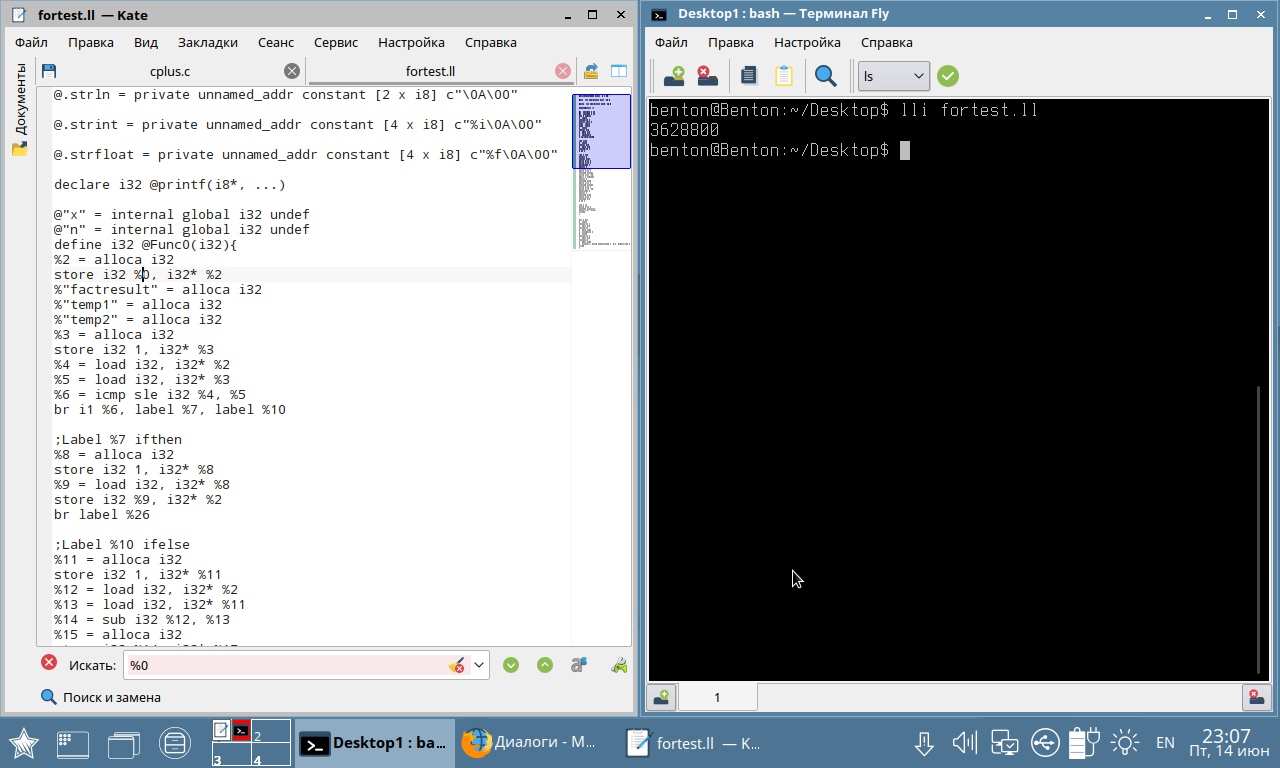


Рисунок 3. Запуск листинга 19

Если входной файл написать с ошибкой, например, присвоить значение переменной, которая не объявлена, компилятор выдаст ошибку. В ошибке будет описана причина ошибки и её местоположение (строка и столбец) во входном файле (рисунок 4).

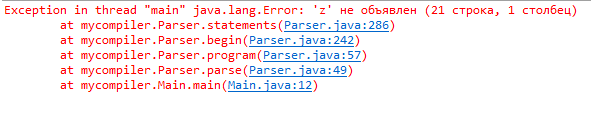


Рисунок 4. Определение ошибок во время парсинга

# ЗАКЛЮЧЕНИЕ

В процессе выполнения курсовой работы была разработана программа, реализующая компилятор заданного подмножества языка Паскаль с незначительными модификациями. Для её разработки использовался язык программирования java, среда программной разработки Eclipse Oxygen.

**Приложение А**

<pascal program> ->

[<program stat>]

<declarations>

<begin-statement>

<program stat> -> E

<type> ->

integer

real

<var> ->

[varname]

<numb> ->

[intlet]

[reallet]

<num> ->

<var>

<numb>

<declarations> ->

<var decl><declarations>

<type \_\_\_\_\_\_,,\_\_\_\_\_\_>

<function \_\_\_\_\_\_,,\_\_\_\_\_\_>

-> E

<function decl> -> function <name> (params) : <type>;

<declarations>

<begin-statement>

<statement> -> <function call>

<var decl> ->

var[<namelist>: <type>;]^+

<begin\_statement> ->

begin <stats> end

<stats> ->

<while stat> ->

while <cond> <begin\_statement>

<if> ->

if <cond> then <begin\_statement>

if <cond> then <begin\_statement> else <begin\_statement>

<for> ->

for <num> to <num> do <begin\_statement>

<writeStat> ->

writeln (<num>)

<function call> ->

[functionname] (<num>,^+)

<assignment>

<var> := <num>

<cond> ->

<num> > <num>

<num> < <num>

<num> >= <num>

<num> <= <num>

<num> <> <num>

<num> and <num>

<num> || <num>