Федеральное государственное автономное образовательное учреждение высшего образования «СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт космических и информационных технологий	
институт	
Кафедра информатики	
кафедра	

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №2

Разработка блока синтаксического и контекстного анализа компилятора

тема

Преподаватель			А. С. Кузнецов
Студент	КИ19-04-1М, 031943329	подпись, дата	инициалы, фамилия И. С. Байкалов
Студент	KIII 3-04-11VI, US 1 343323		и. С. Банкалов
	номер группы, зачетной книжки	подпись, дата	инициалы, фамилия

СОДЕРЖАНИЕ

1 Цель	3
2 Задачи	3
3 Ход работы	3
3.1 Описание семантики основных конструкций языка	3
3.1.1 Объявление переменных	4
3.1.2 Присваивание переменных	5
3.1.3 Условный оператор	5
3.1.4 Циклический оператор	6
3.1.5 Тело условного и циклического оператора	6
3.1.6 Выражения	6
3.2 Таблица приоритетов и ассоциативности операций	7
3.3 Листинг кода полученного анализатора	8
3.4 Примеры работы анализатора	14
4 Вывод	20

1 Цель

Изучение основных методов организации таблиц идентификаторов в компиляторах языков программирования и методов синтаксического и контекстного анализа с их программной реализацией.

2 Задачи

- 1. Изучение теоретического материала об основных методах организации таблиц имен.
- 2. Изучение теоретического материала по организации синтаксического и контекстного анализа языков программирования.
- 3. Составление формального описания процессов синтаксического и контекстного анализа.
 - 4. Программная реализация по формальному описанию.

3 Ход работы

В ходе работы был реализован парсер с использованием Bison, были переписаны функции и типы для построения и вывода AST дерева, а также был переписан код, отвечающий за лексический анализ под вид, необходимый Bison.

3.1 Описание семантики основных конструкций языка

Далее проведем описание синтаксиса и семантики основных конструкций языка.

Сам язык является блочным, то есть состоит из определенных блоков, таких как:

- объявление переменных (declaration);
- присваивание переменным значений (assignment);
- условный оператор IF-ELSE (condition);

- циклический оператор WHILE (loop statement);
- тела условных и циклических операций (compound statement);

В переводе грамматику правило представления в блоках имеет вид:

```
statement:
assignment
| condition
| declaration
| compound_statement
| loop_statement;
```

3.1.1 Объявление переменных

Объявление переменных имеет вид: «type name;», где

«type» - тип данных, к которому принадлежит переменная, «пате» - имя переменной (идентификатор). Тогда, переводя этот блок в грамматику, учитывая, что у нас всего 4 типа данных получаем 4 следующих правила:

- INTEGER TYPE IDENTIFIER SEMICOLON
- FLOAT_TYPE IDENTIFIER SEMICOLON
- STRING_TYPE IDENTIFIER SEMICOLON
- CHAR TYPE IDENTIFIER SEMICOLON

Переменная должна быть обязательно объявлена, если попытаться использовать ее без объявления, то будет выдано сообщение: «Variable x isn't declared».

Переменную нужно объявлять только один раз, при попытке объявить переменную повторно, будет выдано сообщение: «Variable x is already declared».

3.1.2 Присваивание переменных

Присваивание переменных имеет 2 способа. Первый способ: «name = expr;», где «name» - имя переменной (идентификатор) и «expr» - какое-то выражение, которое будет определенно далее. Данное правило в виде грамматики будет иметь вид:

- IDENTIFIER ASSIGN expr SEMICOLON

Второй способ: «type name = expr», то есть присваивание значение сразу при объявлении переменной. Данные правила грамматики имеют вид:

- INTEGER TYPE IDENTIFIER ASSIGN expr SEMICOLON
- FLOAT TYPE IDENTIFIER ASSIGN expr SEMICOLON
- STRING_TYPE IDENTIFIER ASSIGN expr SEMICOLON
- CHAR TYPE IDENTIFIER ASSIGN expr SEMICOLON

Присваивать переменным можно только значения аналогичного типа, если тип отличается, то получим сообщение: «types incompatible in assignment»

3.1.3 Условный оператор

Условный оператор имеет вид: «if (expr) { statement }» или «if (expr) { statement } else { statement }», где «if» и «else» - ключевые слова, «statement» - это тело внутри оператора, которое может содержать математические выражения, присваивания, другие условные и циклические операторы. В виде грамматики будет 2 правила:

- IF OPENPAREN expr CLOSEPAREN statement %prec IFX
- IF OPENPAREN expr CLOSEPAREN statement ELSE statement

Семантика тела условного и циклического оператора описана в пункте 3.1.5.

Семантика условия будет описана в пункте 3.1.6 в виде выражения.

3.1.4 Циклический оператор

Циклический оператор имеет вид: «while (expt) { statement }», где «while» - ключевое слово, «expr» - какое-то выражение и «statement» — это тело оператора. В переводе на грамматику будет иметь вид:

- WHILE OPENPAREN expr CLOSEPAREN statement

Тело циклического оператора не должно быть пустым, в противном случае, будет выдано сообщение: «syntax error, unexpected }».

Семантика тела условного и циклического оператора описана в пункте 3.1.5.

Семантика условия будет описана в пункте 3.1.6 в виде выражения.

3.1.5 Тело условного и циклического оператора

Тела условных и циклических операций имеют вид: { statement_list }, где «statement_list» может состоять из любых блоков программы, определенных ранее, то есть переводя на грамматику, правило будет иметь вид:

- OPENBRACE statement_list CLOSEBRACE

Где «statement list» может содержать в себе множество «statement».

Тело условного или циклического оператора не должно быть пустым, в противном случае, будет выдано сообщение: «syntax error, unexpected }».

3.1.6 Выражения

Еще одна важная конструкция языка — это выражения. Выражения можно представить во множестве видов, например «a = b + 1» или «x = y» и так далее, то есть нужно учитывать, что у нас имеется 4 типа данных (integer, float, string и char), такие операции как «+»,«-», «*», «*», «*», «*», «*», «*», «*», «с» и «==», скобки, а также можно производить операции уже с выражениями, то есть получаем рекурсивное правило для грамматики, правило имеет имя «ехрг», далее представлены правила для выражений:

- expr COMPARE expr;
- expr ADD expr;
- expr SUBSTRACT expr;
- expr MULOPERATOR expr;
- OPENPAREN expr CLOSEPAREN;
- SUBSTRACT expr %prec UMINUS;
- INTEGER_CONST;
- FLOAT CONST;
- STRING CONST;
- CHAR CONST;
- IDENTIFIER.

Сравнивать можно только операнды одного типа, иначе будет выдано сообщение: «error occurred: Comparison error».

Складывать можно только операнды одного типа, при попытке провести операции с операндами разного типа получим ошибку: «error occurred: Adding error», числа складываются, а для строк выполняется операция конкатенации.

Производить операцию вычитания можно только над операндами одного типа иначе получим ошибку: «error occurred: Substract error».

При перемножении или делении значений разных типов мы получим ошибку: «error occurred: Mulop error».

3.2 Таблица приоритетов и ассоциативности операций

В таблице 1 приведены значения приоритетов и ассоциативности операций.

Таблица 1 – Приоритеты и ассоциативности операций

Приоритет	Операции	Описание	Ассоциативность
1	-	Унарный минус	Справа налево

2	+/-/*//	Бинарные сложение, вычитание,	Слева направо
		умножение и деление	
3	/==	Операции сравнения	Слева направо
4	=	Присваивание	Справа налево

3.3 Листинг кода полученного анализатора

Листинг кода полученного синтаксического и контекстного анализатора представлено в таблице 2.

Таблица 2 – Листинг кода парсера.

```
#include <cstdio>
#include <cstdlib>
#include <string>
#include <cstring>
#include <algorithm>
#include <fstream>
#include "ast.hpp"
#include "symtable.hpp"
#include "olmeca driver.hpp"
%skeleton "lalr1.cc"
%require "3.0"
%defines
%define api.parser.class {Olmeca}
%define parse.assert
%code requires {
 class Olmeca_driver;
%param { Olmeca_driver& driver }
%locations
%initial-action {
 @$.begin.filename = @$.end.filename = &driver.filename;
%define parse.trace
%define parse.error verbose
%code {
#undef yyerror
```

```
#define yyerror driver.error
static std::string ErrorMessageVariableNotDeclared(std::string);
static std::string ErrorMessageVariableDoublyDeclared(std::string);
int g LoopNestingCounter = 0;
static TSymbolTable* g TopLevelUserVariableTable =
CreateUserVariableTable(NULL);
static TSymbolTable* currentTable = g TopLevelUserVariableTable;
%union {
 NodeAST* node;
 float fNumber;
 int iNumber;
 std::string* string;
 std::string* identifier;
 char comparison[3];
 char ch[1];
                      EOFILE 0 "end of file"
%token
                      INTEGER TYPE "int"
%token
          <iNumber> INTEGER CONST
                                      "integer"
%token
                     FLOAT TYPE "float type"
%token
          <fNumber> FLOAT CONST
                                      "float"
%token
                      STRING TYPE
%t.oken
                      CHAR TYPE
%t.oken
                      STRING CONST
%t.oken
          <string>
                      CHAR CONST
%token
          <ch>
                      ASSIGN
%token
          <identifier> IDENTIFIER "name"
%token
                      SEMICOLON ";"
%token
                      COLON
%token
                                  " + "
                      ADD
%token
                      SUBSTRACT "-"
%token
                                          "mulop"
          <comparison> MULOPERATOR
%token
                                      "compare"
%token
          <comparison>
                           COMPARE
                      OPENPAREN "("
%token
                      CLOSEPAREN ")"
%token
                      OPENBRACE "{"
%token
                      CLOSEBRACE "}"
%token
                                 "if"
%token
                      ΙF
                                "else"
%token
                      ELSE
%token
                      WHILE
                                       "while"
%token
                      IFX
                      COMMA
                                       ","
%token
%type <node> expr condition assignment statement compound statement
statement list statement list tail declaration loop head loop statement prog
%nonassoc IFX
%nonassoc ELSE
         ASSIGN
%right
%left
          COMPARE
          ADD SUBSTRACT
%left
         MULOPERATOR
%left
%right
          UMINUS
%start prog
%printer { yyoutput << $$; } <*>;
```

```
%destructor { delete $$; } IDENTIFIER
응응
prog:
 statement list {
   if(driver.AST dumping) {
     PrintAST($1, 0);
   driver.result = 0;
  } ;
statement list:
  statement statement list tail {
    if($2 == NULL) {
     $$ = $1;
    } else {
     $$ = CreateNodeAST(typeList, "Stmt List", $1, $2);
statement list tail :
  %empty {
   $$ = NULL;
  | statement list {
   $$ = $1;
  } ;
statement :
 assignment
  | condition
  | declaration
  | compound statement
  | loop_statement;
compound statement :
 OPENBRACE {
   currentTable = CreateUserVariableTable(currentTable);
 statement list
  CLOSEBRACE {
   $$ = $3;
   HideUserVariableTable(currentTable); currentTable = currentTable-
>parentTable;
 } ;
assignment :
  IDENTIFIER ASSIGN expr SEMICOLON {
   TSymbolTableElementPtr var =
LookupUserVariableTableRecursive(currentTable, *$1);
   if (NULL == var) {
     yyerror(ErrorMessageVariableNotDeclared(*$1));
    } else if ($3->valueType != var->table->data[var->index].valueType) {
     yyerror("warning - types incompatible in assignment \n");
    $$ = CreateAssignmentNode(var, $3);
  | INTEGER TYPE IDENTIFIER ASSIGN expr SEMICOLON {
   TSymbolTableElementPtr var =
LookupUserVariableTableRecursive(currentTable, *$2);
   if(var != NULL) {
      yyerror(ErrorMessageVariableDoublyDeclared(*$2));
    } else {
```

```
InsertUserVariableTable(currentTable, *$2, typeInt, var);
    if ($4->valueType != var->table->data[var->index].valueType) {
      yyerror("warning - types incompatible in assignment \n");
    $$ = CreateAssignmentNode(var, $4);
  | FLOAT TYPE IDENTIFIER ASSIGN expr SEMICOLON {
    TSymbolTableElementPtr var =
LookupUserVariableTableRecursive(currentTable, *$2);
   if(var != NULL) {
     yyerror(ErrorMessageVariableDoublyDeclared(*$2));
     InsertUserVariableTable(currentTable, *$2, typeFloat, var);
    if ($4->valueType != var->table->data[var->index].valueType) {
      yyerror("warning - types incompatible in assignment \n");
    $$ = CreateAssignmentNode(var, $4);
  }
  | STRING TYPE IDENTIFIER ASSIGN expr SEMICOLON {
    TSymbolTableElementPtr var =
LookupUserVariableTableRecursive(currentTable, *$2);
   if(var != NULL) {
      yyerror(ErrorMessageVariableDoublyDeclared(*$2));
    } else {
     InsertUserVariableTable(currentTable, *$2, typeString, var);
    if ($4->valueType != var->table->data[var->index].valueType) {
      yyerror("warning - types incompatible in assignment \n");
    $$ = CreateAssignmentNode(var, $4);
  } | CHAR TYPE IDENTIFIER ASSIGN expr SEMICOLON {
    TSymbolTableElementPtr var =
LookupUserVariableTableRecursive(currentTable, *$2);
    if(var != NULL) {
      yyerror(ErrorMessageVariableDoublyDeclared(*$2));
    } else {
      InsertUserVariableTable(currentTable, *$2, typeChar, var);
    if ($4->valueType != var->table->data[var->index].valueType) {
     yyerror("warning - types incompatible in assignment \n");
    $$ = CreateAssignmentNode(var, $4);
declaration :
  INTEGER TYPE IDENTIFIER SEMICOLON {
    TSymbolTableElementPtr var =
LookupUserVariableTableRecursive(currentTable, *$2);
   if(var != NULL) {
     yyerror(ErrorMessageVariableDoublyDeclared(*$2));
    } else {
     InsertUserVariableTable(currentTable, *$2, typeInt, var);
    $$ = CreateAssignmentNode(var, CreateIntegerNode(0));
  | FLOAT TYPE IDENTIFIER SEMICOLON {
    TSymbolTableElementPtr var = LookupUserVariableTable(currentTable, *$2);
    if(var != NULL) {
     yyerror(ErrorMessageVariableDoublyDeclared(*$2));
    } else {
      InsertUserVariableTable(currentTable, *$2, typeFloat, var);
```

```
$$ = CreateAssignmentNode(var, CreateFloatNode(0.0f));
  | STRING TYPE IDENTIFIER SEMICOLON {
   TSymbolTableElementPtr var = LookupUserVariableTable(currentTable, *$2);
   if(var != NULL) {
     yyerror(ErrorMessageVariableDoublyDeclared(*$2));
    } else {
     InsertUserVariableTable(currentTable, *$2, typeString, var);
    $$ = CreateAssignmentNode(var, CreateStringNode(NULL));
  | CHAR TYPE IDENTIFIER SEMICOLON {
   TSymbolTableElementPtr var = LookupUserVariableTable(currentTable, *$2);
   if(var != NULL) {
     yyerror(ErrorMessageVariableDoublyDeclared(*$2));
    } else {
     InsertUserVariableTable(currentTable, *$2, typeChar, var);
   $$ = CreateAssignmentNode(var, CreateCharNode(NULL));
  };
condition :
 IF OPENPAREN expr CLOSEPAREN statement %prec IFX {
   $$ = CreateControlFlowNode(typeIfStatement, $3, $5, NULL);
  | IF OPENPAREN expr CLOSEPAREN statement ELSE statement {
   $$ = CreateControlFlowNode(typeIfStatement, $3, $5, $7);
  };
loop statement :
 loop head statement {
   $$ = CreateControlFlowNode(typeWhileStatement, $1, $2, NULL);
    --g LoopNestingCounter;
  };
loop head :
 WHILE OPENPAREN expr CLOSEPAREN {
   $$ = $3;
   ++g LoopNestingCounter;
expr :
 expr COMPARE expr {
   if($1->valueType != $3->valueType) {
      yyerror("Cannot compare different types \n");
      $$ = CreateErrorNode("Comparison error \n");
    } else {
      $$ = CreateNodeAST(typeBinaryOp, $2, $1, $3);
   }
  }
  | expr ADD expr {
   if($1->valueType != $3->valueType) {
     yyerror("Cannot add different types \n");
     $$ = CreateErrorNode("Adding error \n");
    } else {
     $$ = CreateNodeAST(typeBinaryOp, "+", $1, $3);
    }
  | expr SUBSTRACT expr {
   if($1->valueType != $3->valueType) {
     yyerror("Cannot substract different types \n");
      $$ = CreateErrorNode("Substract error \n");
```

```
} else if ($1->valueType == typeString || $3->valueType == typeString) {
      yyerror("Cannot substract string types \n");
      $$ = CreateErrorNode("Substract error \n");
    } else {
      $$ = CreateNodeAST(typeBinaryOp, "-", $1, $3);
  | expr MULOPERATOR expr {
   if($1->valueType != $3->valueType) {
      yyerror("Cannot mulop different types \n");
      $$ = CreateErrorNode("Mulop error \n");
    } else if ($1->valueType == typeString || $3->valueType == typeString) {
      yyerror("Cannot substract string types \n");
      $$ = CreateErrorNode("Substract error \n");
      $$ = CreateNodeAST(typeBinaryOp, $2, $1, $3);
  }
  | OPENPAREN expr CLOSEPAREN {
   $$ = $2;
  | SUBSTRACT expr %prec UMINUS {
   if ($2->valueType == typeString) {
      yyerror("Cannot unary substract string types \n");
      $$ = CreateErrorNode("Unary substract error \n");
    } else {
      $$ = CreateNodeAST(typeUnaryOp, "-", $2, NULL);
  | INTEGER CONST {
   $$ = CreateIntegerNode($1);
  | FLOAT CONST {
   $$ = \overline{CreateFloatNode($1)};
  | STRING CONST {
   $$ = CreateStringNode($1);
  | CHAR CONST {
   $$ = CreateCharNode($1);
  | IDENTIFIER {
   TSymbolTableElementPtr var =
LookupUserVariableTableRecursive(currentTable, *$1);
   if (NULL == var) {
     yyerror(ErrorMessageVariableNotDeclared(*$1));
   $$ = CreateReferenceNode(var);
 };
void yy::Olmeca::error(const location type& 1, const std::string& m) {
 driver.error(l, m);
static std::string ErrorMessageVariableNotDeclared(std::string name) {
 std::string errorDeclaration = "error - Variable " + name + " isn't
declared";
 return errorDeclaration;
static std::string ErrorMessageVariableDoublyDeclared(std::string name) {
```

```
std::string errorDeclaration = "error - Variable " + name + " is already
declared";
  return errorDeclaration;
}
```

3.4 Примеры работы анализатора

Разработанный парсер умеет диагностировать такие ошибки, как:

- лексические ошибки символы и токены, не входящие в состав языка;
- синтаксические ошибки неверные последовательности токенов/симво лов, входящих в язык;
- семантические ошибки повторное определение переменных, использо вание неопределенных переменных, арифметические операции, операции сравнения и присвоения между разными типами.

Пример диагностики ошибок можно увидеть на рисунках 1 - 5.

Рисунок 1 – Символы, не входящие в состав языка. Лексическая ошибка

```
1    string a = "test;
2

Terminal: Local × +

EnotBoris:lab2 ilyabaikalow$ ./olmeca tests/11.olm
tests/11.olm: 1.12-17: unfinished string literal "test;
tests/11.olm: tests/11.olm:1.1: syntax error, unexpected invalid token
EnotBoris:lab2 ilyabaikalow$
```

Рисунок 2 — Строка не закрыта вторым знаком двойных кавычек. Лексическая ощибка

```
1    string a = "test1";
2    string a = "test2";
3

Terminal: Local × +

EnotBoris:lab2 ilyabaikalow$ ./olmeca tests/11.olm
tests/11.olm: 2.19: error - Variable a is already declared
variable a =
    string "test1"
    a =
    string "test2"
Olmeca is good!
```

Рисунок 3 – Повторное объявление переменной. Семантическая ошибка

```
₫ 11.olm ×
      integer a = 1;
     float b = 1.2;
     if (a > b) {
          a = 0;
5
      }
Terminal: Local ×
                  +
EnotBoris:lab2 ilyabaikalow$ ./olmeca tests/11.olm
tests/11.olm: 4.10: Cannot compare different types
variable a =
integer 1
variable b =
 float 1.2
 if
 error occurred: Comparison error
   true way
  a =
  integer 0
Olmeca is good!
```

Рисунок 4 – Сравнение переменных разного типа. Семантическая ошибка

Рисунок 5 – Использование необъявленной переменной. Семантическая ошибка

Парсер обеспечивает распознавание всех грамматических конструкций языка определенных в «белой книге» из первой самостоятельной работы.

Остальные примеры входных файлов и вывод для них AST дерева представлены на рисунках 6-10.

Рисунок 6 – Пример вывода дерева AST

```
| Second | S
```

Рисунок 7 – Пример вывода дерева AST

```
| March | Marc
```

Рисунок 8 – Пример вывода дерева AST

```
| Description | March | March
```

Рисунок 9 – Пример вывода дерева AST

```
integer a;
a = 12;

integer b = 10;

Terminal: Local × +

EnotBoris:lab2 ilyabaikalow$ ./olmeca tests/1.olm
variable a =
integer 0
variable a =
integer 12
b =
integer 10
Olmeca is good!
```

Рисунок 10 – Пример вывода дерева AST

4 Вывод

В ходе практической работы был реализован синтаксический анализатор с использованием Bison. Были изучены основы построения таблицы идентификаторов и абстрактного синтаксического дерева при синтаксическом разборе языков программирования.