МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования «БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ

ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Факультет Информационных Технологий

Кафедра Программной инженерии

Специальность 1-40 01 01 Программное обеспечение информационных технологий

Специализация Программирование интернет-приложений

**ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

**К КУРСОВОМУ ПРОЕКТУ НА ТЕМУ:**

«Разработка компилятора VAD-2021»

Выполнил студент Валдайцев Александр Денисович

(Ф.И.О.)

Руководитель проекта ст.пр. Наркевич Аделина Сергеевна

(учен. степень, звание, должность, подпись, Ф.И.О.)

Заведующий кафедрой к.т.н. доц. Пацей Наталья Владимировна

(учен. степень, звание, должность, подпись, Ф.И.О.)

Консультант асс. Мущук Артур Николаевич

(учен. степень, звание, должность, подпись, Ф.И.О.)

Нормоконтролер асс. Мущук Артур Николаевич

(учен. степень, звание, должность, подпись, Ф.И.О.)

Курсовой проект защищен с оценкой

Минск 2021

Оглавление

[Введение 4](#_Toc532650587)

[1 Спецификация языка программирования 5](#_Toc532650588)

[1.1 Характеристика языка программирования 5](#_Toc532650589)

[1.2 Определение алфавита языка программирования 5](#_Toc532650590)

[1.3 Применяемые сепараторы 5](#_Toc532650591)

[1.4 Применяемые кодировки 6](#_Toc532650592)

[1.5 Типы данных 6](#_Toc532650593)

[1.6 Преобразование типов данных 7](#_Toc532650594)

[1.7 Идентификаторы 7](#_Toc532650595)

[1.8 Литералы 7](#_Toc532650596)

[1.9 Объявление данных 8](#_Toc532650597)

[1.10 Инициализация данных 8](#_Toc532650598)

[1.11 Инструкции языка 8](#_Toc532650599)

[1.12 Операции языка 9](#_Toc532650600)

[1.13 Выражения и их вычисление 9](#_Toc532650601)

[1.14 Конструкции языка 10](#_Toc532650602)

[1.15 Области видимости идентификаторов 11](#_Toc532650603)

[1.16 Семантические проверки 11](#_Toc532650604)

[1.17 Распределение оперативной памяти на этапе выполнения 11](#_Toc532650605)

[1.18 Стандартная библиотека и её состав 12](#_Toc532650606)

[1.19 Ввод и вывод данных 12](#_Toc532650607)

[1.20 Точка входа 12](#_Toc532650608)

[1.21 Препроцессор 13](#_Toc532650609)

[1.22 Соглашения о вызовах 13](#_Toc532650610)

[1.23 Объектный код 13](#_Toc532650611)

[1.24 Классификация сообщений транслятора 13](#_Toc532650612)

[1.25 Контрольный пример 13](#_Toc532650613)

[2 Структура транслятора 14](#_Toc532650614)

[2.1 Компоненты транслятора, их назначение и принципы взаимодействия 14](#_Toc532650615)

[2.2 Перечень входных параметров транслятора 15](#_Toc532650616)

[2.3 Перечень протоколов, формируемых транслятором и их содержимое 15](#_Toc532650617)

[3 Разработка лексического анализатора 17](#_Toc532650618)

[3.1 Структура лексического анализатора 17](#_Toc532650619)

[3.2. Контроль входных символов 18](#_Toc532650620)

[3.3 Удаление избыточных символов 18](#_Toc532650621)

[3.4 Перечень ключевых слов 19](#_Toc532650622)

[3.5 Основные структуры данных 20](#_Toc532650623)

[3.6 Принцип обработки ошибок 22](#_Toc532650624)

[3.7 Структура и перечень сообщений лексического анализатора 22](#_Toc532650625)

[3.8 Параметры лексического анализатора 22](#_Toc532650626)

[3.9 Алгоритм лексического анализа 22](#_Toc532650627)

[3.10 Контрольный пример 23](#_Toc532650628)

[4. Разработка синтаксического анализатора 24](#_Toc532650629)

[4.1 Структура синтаксического анализатора 24](#_Toc532650630)

[4.2 Контекстно-свободная грамматика, описывающая синтаксис языка 24](#_Toc532650631)

[4.3 Построение конечного магазинного автомата 26](#_Toc532650632)

[4.4 Основные структуры данных 27](#_Toc532650633)

[4.5 Описание алгоритма синтаксического разбора 27](#_Toc532650634)

[4.6 Структура и перечень сообщений синтаксического анализатора 28](#_Toc532650635)

[4.7. Параметры синтаксического анализатора и режимы его работы 28](#_Toc532650636)

[4.8. Принцип обработки ошибок 28](#_Toc532650637)

[4.9. Контрольный пример 28](#_Toc532650638)

[5 Разработка семантического анализатора 29](#_Toc532650639)

[5.1 Структура семантического анализатора 29](#_Toc532650640)

[5.2 Функции семантического анализатора 29](#_Toc532650641)

[5.3 Структура и перечень сообщений семантического анализатора 29](#_Toc532650642)

[5.4 Принцип обработки ошибок 30](#_Toc532650643)

[5.5 Контрольный пример 30](#_Toc532650644)

[6. Вычисление выражений 31](#_Toc532650645)

[6.1 Выражения, допускаемые языком 31](#_Toc532650646)

[6.2 Польская запись и принцип её построения 31](#_Toc532650647)

[6.3 Программная реализация обработки выражений 31](#_Toc532650648)

[6.4 Контрольный пример 32](#_Toc532650649)

[7. Генерация кода 33](#_Toc532650650)

[7.1 Структура генератора кода 33](#_Toc532650651)

[7.2 Представление типов данных в оперативной памяти 33](#_Toc532650652)

[7.3 Статическая библиотека 34](#_Toc532650653)

[7.4 Особенности алгоритма генерации кода 34](#_Toc532650654)

[7.5 Входные параметры генератора кода 34](#_Toc532650655)

[7.6 Контрольный пример 34](#_Toc532650656)

[8. Тестирование транслятора 36](#_Toc532650657)

[8.1 Тестирование проверки на допустимость символов 36](#_Toc532650658)

[8.2 Тестирование лексического анализатора 36](#_Toc532650659)

[8.3 Тестирование синтаксического анализатора 36](#_Toc532650660)

[8.4 Тестирование семантического анализатора 37](#_Toc532650661)

[Заключение 39](#_Toc532650662)

[Список использованных источников 40](#_Toc532650663)

[Приложение А 41](#_Toc532650664)

[Приложение Б 42](#_Toc532650665)

[Приложение В 46](#_Toc532650666)

[Приложение Г 51](#_Toc532650667)

[Приложение Д 55](#_Toc532650668)

# Введение

В данном курсовом проекте поставлена задача разработки собственного языка программирования и транслятора для него. Название языка – VAD-2021. Написание транслятора будет осуществляться на языке C++, при этом код на языке VAD-2021 будет транслироваться в язык ассемблера.

Задание на курсовой проект можно разделить на следующие задачи:

1. Разработка спецификации языка VAD-2021;
2. Разработка лексического анализатора;
3. Разработка синтаксического анализатора;
4. Разработка семантического анализатора;
5. Разбор арифметических выражений;
6. Разработка генератора кода;
7. Тестирование транслятора.

# 1 Спецификация языка программирования

## 1.1 Характеристика языка программирования

Язык программирования VAD-2021 является процедурным, универсальным, строго типизированным, не объектно-ориентированным, компилируемым.

## 1.2 Определение алфавита языка программирования

При написании программы на языке VAD-2021 используется таблица символов ASCII, представленная в рис.1.1.

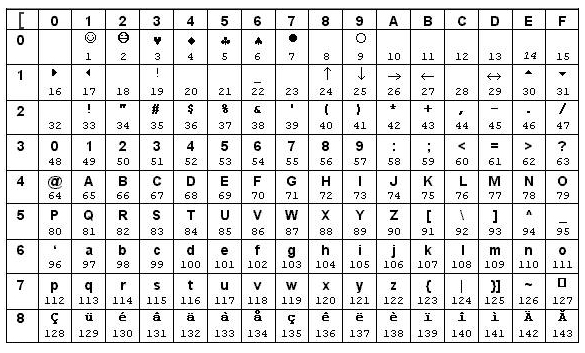


Рисунок 1.1 Алфавит входных символов

Символы, используемые на этапе выполнения: [a…z], [A…Z], [0…9], символы пробела, табуляции и перевода строки, спецсимволы: [ ] ( ) , ; : # + - / \* % > < & !.

## 1.3 Применяемые сепараторы

Символы-сепараторы служат в качестве разделителей цепочек языка во время обработки исходного текста программы с целью разделения на токены. Они представлены в таблице 1.1.

Таблица 1.1 Символы-сепараторы

|  |  |
| --- | --- |
| Символ(ы) | Назначение |
| ‘пробел’ | Разделитель цепочек. Допускается везде кроме названий идентификаторов и ключевых слов |

Продолжение таблицы 1.1

|  |  |
| --- | --- |
| Символ(ы) | Назначение |
| **[** … **]** | Блок функции или условной конструкции/цикла |
| **(** … **)** | Блок фактических или формальных параметров функции, а также приоритет арифметических операций |
| **,** | Разделитель параметров функций |
| **#** | Символ, отделяющий условные конструкции/циклы |
| **+ - \* /** % | Арифметические операции |
| **> <** & ! | Логические операции (операции сравнения: больше, меньше, проверка на равенство, на неравенство), используемые в условии цикла/условной конструкции. |
| ; | Разделитель программных конструкций |
| }{ | Операторы сдвигов |
| **=** | Оператор присваивания |

## 1.4 Применяемые кодировки

Для написания программ язык VAD-2021 использует кодировку ASCII, содержащую английский алфавит, а также некоторые специальные символы, такие как [ ] ( ) , ; : # + - / \* % > < & !{}.

## 1.5 Типы данных

В языке VAD-2021 реализованы два фундаментальных типа данных: целочисленный и строковый. Описание типов приведено в таблице 1.2.

Таблица 1.2 Типы данных языка VAD-2021

|  |  |
| --- | --- |
| Тип данных | Характеристика |
| Целочисленный тип данных **number** | Фундаментальный тип данных. Используется для работы с числовыми значениями. В памяти занимает 2 байта. Максимальное значение: 255.  Минимальное значение: -256.  Инициализация по умолчанию: значение 0.  Поддерживаемые операции:  **+** (бинарный) – оператор сложения;  - (бинарный) – оператор вычитания;  **\*** (бинарный) – оператор умножения;  **/** (бинарный) – оператор деления нацело;  % (бинарный) – оператор остатка от деления;  **=** (бинарный) – оператор присваивания.  В качестве оператора цикла/условия поддерживаются операции:  **>** (бинарный) – оператор «больше»;  **<** (бинарный) – оператор «меньше»;  ! (бинарный) – оператор проверки на неравенство. |

Продолжение таблицы 1.2

|  |  |
| --- | --- |
| Тип данных | Характеристика |
| Строковый тип данных **string** | Фундаментальный тип данных. Используется для работы с символами, каждый из которых занимает 1 байт. Максимальное количество символов – 256.  Инициализация по умолчанию: строка нулевой длины “”. Операции над данными строкового типа: присваивание строковому идентификатору значения другого строкового идентификатора, строкового литерала или значения строковой функции, а также использование библиотечных функций. |

## 1.6 Преобразование типов данных

Преобразование типов данных в языке VAD-2021 не предусмотрено.

## 1.7 Идентификаторы

Общее количество идентификаторов ограничено максимальным размером таблицы идентификаторов. Идентификаторы должны содержать только символы нижнего регистра латинского алфавита. Максимальная длина идентификатора равна восьми символам. Идентификаторы, объявленные внутри функционального блока, получают префикс, идентичный имени функции, внутри которой они объявлены. Префикс занимает 8 дополнительных символов. В случае превышения заданной длины, идентификаторы усекаются до длины, равной 16 символов (8 символов на имя идентификатора, 8 символов на префикс). Данные правила действуют для всех типов идентификаторов. Зарезервированные идентификаторы не предусмотрены. Идентификаторы не должны совпадать с ключевыми словами. Типы идентификаторов: имя переменной, имя функции, параметр функции. Имена идентификаторов-функций не должны совпадать с именами команд ассемблера (это не касается имён идентификаторов-переменных).

## 1.8 Литералы

С помощью литералов осуществляется инициализация переменных. Все литералы являются **rvalue**. Типы литералов языка VAD-2021 представлены в таблице 1.3.

Таблица 1.3 Литералы

|  |  |
| --- | --- |
| Литералы | Пояснение |
| Целочисленные литералы в десятичном представлении | Последовательность цифр 0…9 с предшествующим знаком минус или без него (знак минус не должен отделяться пробелом) |
| Строковые литералы | Набор символов (от 1 до 255), заключённых в двойные кавычки |

Продолжение таблицы 1.3

|  |  |
| --- | --- |
| Целочисленные литералы в шестнадцатеричном представлении | Последовательность цифр 0…9 и символов A…F с предшествующим символом ‘q’ |

Ограничения на строковые литералы языка VAD-2021: внутри литерала не допускается использование символов кириллицы, а также одинарных и двойных кавычек. Ограничения на целочисленные литералы: не могут начинаться с нуля, если их значение не ноль; если литерал отрицательный, после знака “-” не может быть нуля.

## 1.9 Объявление данных

Для объявления переменной используется ключевое слово **new**, после которого указывается тип данных и имя идентификатора. Допускается инициализация при объявлении.

Пример объявления числового типа с инициализацией:

**new** **number** num1 = -1

**new** **number** num2 = qC

Пример объявления переменной символьного типа с инициализацией:

**new** **string** str1= “hello world”

Для объявления функций используется ключевое слово **function**, перед которым указывается тип функции (если функция возвращает значение), или ключевое слово **procedure**, если функция ничего не возвращает, а после –имя функции либо скрипта. Далее обязателен список параметров и тело функции.

## 1.10 Инициализация данных

При объявлении переменной допускается инициализация данных. При этом переменной будет присвоено значение литерала или идентификатора, стоящего справа от знака равенства. Объектами-инициализаторами могут быть только идентификаторы или литералы. При объявлении без инициализации предусмотрены значения по умолчанию: значение 0 для типа **number** и строка длины 0 (“”) для типа **string**.

## 1.11 Инструкции языка

Инструкции языка VAD-2021 представлены в таблице 1.4.

Таблица 1.4 ИнструкцииязыкаVAD-2021

|  |  |
| --- | --- |
| Инструкция | Реализация |
| Объявление переменной | **new** <тип данных> <идентификатор>; |

Продолжение таблицы 1.4

|  |  |
| --- | --- |
| Инструкция | Реализация |
| Объявление переменной с явной инициализацией | **new** <тип данных> <идентификатор> = <значение>;  Значение – инициализатор конкретного типа. Может быть только литералом или идентификатором |
| Возврат из функции или скрипта | Для функций, возвращающих значение:  **return** <идентификатор/литерал>;  Для скриптов:  **return;** |
| Вывод данных | **print** <идентификатор/литерал>; |
| Вывод данных | **read** <идентификатор >; |
| Вызов функции или скрипта | <идентификатор функции> (<список параметров>);  Список параметров может быть пустым. |
| Перевод строки | **newline**; |
| Присваивание | <идентификатор> **=** <выражение>;  Выражением может быть идентификатор, литерал, или вызов функции. Для целочисленного типа выражение может быть дополнено арифметическими операциями с любым количеством операндов с использованием скобок. Для строкового типа выражение может быть только идентификатором, литералом или вызовом функции, возвращающей значение строкового типа. |

## 1.12 Операции языка

В языке VAD-2021 предусмотрены следующие операции с данными. Приоритетность операции умножения выше приоритета операций сложения и вычитания. Для установки наивысшего приоритета используются круглые скобки. Операции языка представлены в таблице 1.5.

Таблица 1.5 Операции языка VAD-2021

|  |  |
| --- | --- |
| Тип оператора | Оператор |
| Арифметические | 1. **+ –** сложение 2. - – вычитание 3. \* – умножение 4. **/** – деление нацело 5. % – остаток от деления 6. = – присваивание |
| Строковые | 1. **=** – присваивание |
| Логические | 1. **>** – больше  2. **<** – меньше  3. ! – проверка на неравенство |
| Сдвиговые | 1. } – сдвиг вправо  2. { – сдвиг влево |

## 1.13 Выражения и их вычисление

Вычисление выражений – одна из важнейших задач языков программирования. Всякое выражение составляется согласно следующим правилам:

1. Допускается использовать скобки для смены приоритета операций;
2. Выражение записывается в строку без переносов;
3. Использование двух подряд идущих операторов не допускается;
4. Допускается использовать в выражении вызов функции, вычисляющей и возвращающей целочисленное значение.

Перед генерацией кода каждое выражение приводится к записи в польской записи для удобства дальнейшего вычисления выражения на языке ассемблера.

## 1.14 Конструкции языка

Программа на языке VAD-2021 оформляется в виде функций пользователя и главной функции. При составлении функций рекомендуется выделять блоки и фрагменты и применять отступы для лучшей читаемости кода.

Программные конструкции языка представлены в таблице 1.6.

Таблица 1.6 Программные конструкции языка VAD-2021

|  |  |  |
| --- | --- | --- |
| Конструкция | | Реализация |
| Главная функция | | **main**  **[**  …  **]** |
| Внешняя функция | | <тип данных> **function** <идентификатор> **(**<тип> <идентификатор>, ...**)**  **[**…  **return** <идентификатор/литерал>;  **]** |
| Внешний скрипт | **script** **function** <идентификатор> **(** <тип> <идентификатор>, ...**)**  **[**…  **return**;  **]** | |
| Цикл | **criterion:** <идентификатор1> <оператор> <идентификатор2> **#**  **cycle** **[** … **]** **#**  Цикл (операторы внутри блока **cycle**) выполняется, пока истинно условие “<идентификатор1> <оператор> <идентификатор2>”, имеющее тот же смысл, что и в примере выше. | |

Продолжение таблицы 1.6

|  |  |
| --- | --- |
| Конструкция | Реализация |

|  |  |
| --- | --- |
| Условная конструкция | **criterion:** <идентификатор1> <оператор> <идентификатор2> **#**  **istrue** **[** … **]**  **isfalse** **[** … **]** **#**  <идентификатор1>, <идентификатор2> - идентификаторы или литералы целочисленного типа (но не два литерала одновременно). <оператор> - один из операторов сравнения ( **> < & !** ), устанавливающий отношение между двумя операндами и организующий условие данной конструкции. При истинности условия выполняется код внутри блока **istrue**, иначе – код внутри блока **isfalse**. Любой из блоков **istrue**, **isfalse** может отсутствовать, но не оба блока одновременно. При отсутствии одного из блоков, в зависимости от истинности или ложности условия программа может как выполнить один из заявленных блоков, так и передать управление инструкции, следующей в коде за закрывающим условную конструкцию символом **‘#’**. |

|  |  |
| --- | --- |
| Внешний скрипт | **script function** <идентификатор> **(** <тип> <идентификатор>, ...**)**  **[**  …  **return**;  **]** |

## 1.15 Области видимости идентификаторов

Область видимости: сверху вниз (как и в С++). Переменные, объявленные в одной функции, недоступны в другой. Все объявления и операции с переменными происходят внутри какого-либо блока. Каждая переменная или параметр функции получают префикс – название функции, внутри которой они находятся.

Все идентификаторы являются локальными и обязаны быть объявленными внутри какой-либо функции. Глобальных переменных нет. Параметры видны только внутри функции, в которой объявлены.

## 1.16 Семантические проверки

В языке программирования VAD-2021 выполняются следующие семантические проверки:

1. Наличие функции **main** – точки входа в программу;
2. Единственность точки входа;
3. Переопределение идентификаторов;
4. Использование идентификаторов без их объявления;
5. Проверка соответствия типа функции и возвращаемого параметра;
6. Правильность передаваемых в функцию параметров: количество, типы;
7. Правильность строковых выражений;
8. Превышение размера строковых и числовых литералов;
9. Правильность составленного условия цикла/условного оператора.

## 1.17 Распределение оперативной памяти на этапе выполнения

Транслированный код использует две области памяти. В сегмент констант заносятся все литералы. В сегмент данных заносятся переменные и параметры функций. Локальная область видимости в исходном коде определяется за счет использования правил именования идентификаторов и регулируется их префиксами, что и обуславливает их локальность на уровне исходного кода, несмотря на то, что в оттранслированным в язык ассемблера коде переменные имеют глобальную область видимости.

## 1.18 Стандартная библиотека и её состав

В языке VAD-2021 присутствует стандартная библиотека, которая подключается автоматически при трансляции исходного кода в язык ассемблера. Содержимое библиотеки и описание функций представлено в таблице 1.8.

Таблица 1.8 Стандартная библиотека языка VAD-2021

|  |  |
| --- | --- |
| Функция | Описание |
| **string** strcopy(**string** str); | Строковая функция, выполняет копирование строки str. Возвращает скопированную строку. |
| **number** atoii(**string** str); | Целочисленная функция. Преобразует строку в число |
| **number** length(**string** str); | Целочисленная функция. Вычисляет и возвращает длину строки str |

Стандартная библиотека написана на языке С++, подключается к транслированному коду на этапе генерации кода. Вызовы стандартных функций доступны там же, где и вызов пользовательских функций. Также в стандартной библиотеке реализованы функции для манипулирования выводом, недоступные конечному пользователю. Для вывода предусмотрен оператор **print**. Эти функции представлены в таблице 1.9.

Таблица 1.9 Дополнительные функции стандартной библиотеки

|  |  |
| --- | --- |
| Функция на языке С++ | Описание |
| void outnum(int value) | Функция для вывода в стандартный поток значения целочисленного идентификатора/литерала |
| void outstr(char\* line) | Функция для вывода в стандартный поток значения строкового идентификатора/литерала |

## 1.19 Ввод и вывод данных

Вывод данных осуществляется с помощью оператора **print**. Допускается использование оператора **print** с литералами и идентификаторами.

Функции, управляющие выводом данных, реализованы на языке С++ и вызываются из транслированного кода, конечному пользователю недоступны. Пользовательская команда **print** в транслированном коде будут заменена вызовом нужных библиотечных функций. Библиотека, содержащая нужные скрипты, подключается на этапе генерации кода.

## 1.20 Точка входа

В языке VAD-2021 каждая программа должна содержать главную функцию (точку входа) **main**, с первой инструкции которой начнётся последовательное выполнение команд программы.

## 1.21 Препроцессор

Препроцессор, принимающий и выдающий некоторые данные на вход транслятору, в языке VAD-2021 отсутствует.

## 1.22 Соглашения о вызовах

В языке вызов функций происходит по соглашению о вызовах stdcall. Осо-бенности stdcall:

– все параметры функции передаются через стек;

– память высвобождает вызываемый код;

– занесение в стек параметров идёт справа налево.

## 1.23 Объектный код

Язык VAD-2021 транслируется в язык ассемблера, а затем – в объектный код.

## 1.24 Классификация сообщений транслятора

Генерируемые транслятором сообщения определяют степень его информативности, то есть сообщения транслятора должны давать максимально полную информацию о допущенной пользователем ошибке при написании программы. Сообщения транслятора приведены в таблице 1.10, а также в приложении А.

Таблица 1.10 Классификация ошибок

|  |  |
| --- | --- |
| Номера ошибок | Характеристика |
| 0 – 200 | Системные ошибки |
| 200 – 299 | Ошибки лексического анализа |
| 300 – 399 | Ошибки семантического анализа |
| 600 – 699 | Ошибки синтаксического анализа |
| 400-499, 700-999 | Зарезервированные коды ошибок |

## 1.25 Контрольный пример

Контрольный пример демонстрирует главные особенности языка VAD-2021: его фундаментальные типы, основные структуры, функции, скрипты, использование функций статической библиотеки. Исходный код контрольного примера представлен в приложении А.

# 2 Структура транслятора

## 2.1 Компоненты транслятора, их назначение и принципы взаимодействия

В языке VAD-2021 исходный код транслируется в язык Assembler. Транслятор языка разделён на отдельные части, которые взаимодействуют между собой и выполняют отведённые им функции, которые представлены в пункте 2.1. Для того чтобы получить ассемблерный код, используется выходные данные работы лексического анализатора, а именно таблица лексем и таблица идентификаторов. Для указания выходных файлов используются входные параметры транслятора, которые описаны в таблице 2.1. Структура транслятора языка VAD-2021 приведена на рисунке 1.



Рисунок 2.1 Структура транслятора языка программирования VAD-2021

Первая стадия работы компилятора называется лексическим анализом, а программа, её реализующая, – лексическим анализатором (сканером). На вход лексического анализатора подаётся последовательность символов входного языка. Он производит предварительный разбор текста, преобразующий единый массив текстовых символов в массив отдельных слов (в теории компиляции вместо термина «слово» часто используют термин «токен»). Примеры лексических единиц: идентификаторы, числа, символы операций, служебные слова и т.д. Лексический анализатор преобразует исходный текст, заменяя лексические единицы их внутренним представлением – лексемами, для создания промежуточного представления исходной программы. Каждой лексеме сопоставляется ее тип и запись в таблице идентификаторов, в которой хранится дополнительная информация. Таблица лексем (ТЛ) и таблица идентификаторов (ТИ) являются входом для следующей фазы компилятора – синтаксического анализа (разбора, парсера).

Цели лексического анализатора:

− убрать все лишние пробелы;

− выполнить распознавание лексем;

− построить таблицу лексем и таблицу идентификаторов;

− при неуспешном распознавании или обнаружении некоторых ошибок во входном тексте выдать сообщение об ошибке.

Синтаксический анализатор – часть компилятора, выполняющая синтаксический анализ, то есть проверку исходного кода на соответствие правилам грамматики. Входной информацией для синтаксического анализа является таблица лексем и таблица идентификаторов. Выходной информацией является дерево разбора.

Семантический анализатор – часть транслятора, выполняющая семантический анализ, то есть проверку исходного кода на наличие ошибок, которые невозможно отследить при помощи регулярной и контекстно-свободной грамматики. Входными данными являются таблица лексем и идентификаторов.

Генератор кода – часть транслятора, выполняющая генерацию ассемблерного кода на основе полученных данных на предыдущих этапах трансляции. На вход генератора подаются таблица лексем и таблица идентификаторов, на основе которых генерируется файл с ассемблерным кодом.

## 2.2 Перечень входных параметров транслятора

Для формирования файлов с результатами работы лексического, синтаксического и семантического анализаторов используются входные параметры транслятора, которые приведены в таблице 2.1.

Таблица 2.1 Входные параметры транслятора языка VAD-2021

|  |  |  |
| --- | --- | --- |
| Входной параметр | Описание параметра | Значение по умолчанию |
| -in:<путь к in-файлу> | Файл с исходным кодом на языке VAD-2021 , имеющий расширение .txt | Не предусмотрено |
| -log:<путь к log-файлу> | Файл журнала для вывода протоколов работы программы. | Значение по умолчанию:  <имя in-файла>.log |

## 2.3 Перечень протоколов, формируемых транслятором и их содержимое

В ходе работы программы формируются протоколы работы лексического, синтаксического и семантического анализаторов, которые содержат в себе перечень протоколов работы. В таблице 2.2 приведены протоколы, формируемые транслятором и их содержимое.

Таблица 2.2 Протоколы, формируемые транслятором языка VAD-2021

|  |  |
| --- | --- |
| Формируемый протокол | Описание выходного протокола |
| Файл журнала, заданный параметром "-log:" | Файл с протоколом работы транслятора языка программирования VAD-2021 . Содержит таблицу лексем и таблицу идентификаторов, протокол работы синтаксического анализатора и дерево разбора, полученные на этапе лексического и синтаксического анализа, а также результат работы алгоритма преобразования выражений к польской записи. |
| Выходной файл, c расширением ".asm" | Результат работы программы – файл, содержащий исходный код на языке ассемблера. |

# 3 Разработка лексического анализатора

## 3.1 Структура лексического анализатора

Первая стадия работы компилятора называется лексическим анализом, а программа, её реализующая, – лексическим анализатором (сканером). На вход лексического анализатора подаётся исходный код входного языка. Лексический анализатор выделяет в этой последовательности простейшие конструкции языка,. Лексический анализатор производит предварительный разбор текста, преобразующий единый массив текстовых символов в массив токенов.

Примеры лексических единиц: идентификаторы, числа, символы операций, служебные слова и т.д. Лексический анализатор преобразует исходный текст, заменяя лексические единицы их внутренним представлением – лексемами, для создания промежуточного представления исходной программы. Каждой лексеме сопоставляется ее тип и запись в таблице идентификаторов, в которой хранится дополнительная информация.

Функции лексического анализатора:

− удаление «пустых» символов и комментариев. Если «пустые» символы (пробелы, знаки табуляции и перехода на новую строку) и комментарии будут удалены лексическим анализатором, синтаксический анализатор никогда не столкнется с ними (альтернативный способ, состоящий в модификации грамматики для включения «пустых» символов и комментариев в синтаксис, достаточно сложен для реализации);

− распознавание идентификаторов и ключевых слов;

− распознавание констант;

− распознавание разделителей и знаков операций.

Исходный код программы представлен в приложении А, структура лексического анализатора представлена на рисунке 3.1.



Рисунок 3.1 Структура лексического анализатора

## 3.2. Контроль входных символов

Для удобной работы с исходным кодом, при передаче его в лексический анализатор, все символы разделяются по категориям. Таблица входных символов представлена на рисунке 3.2, категории входных символов представлены в таблице 3.1.



Рисунок 3.2. Таблица контроля входных символов

Таблица 3.1 Соответствие символов и их значений в таблице

|  |  |
| --- | --- |
| Значение в таблице входных символов | Символы |
| Разрешенный | T |
| Запрещенный | F |
| Игнорируемый | I |
| Литерал | Q |
| Сепаратор | S |
| Перевод строки | N |
| Пробел, табуляция | P |

## 3.3 Удаление избыточных символов

Избыточными символами являются символы табуляции и пробелы.

Избыточные символы удаляются на этапе разбиения исходного кода на токены.

Описание алгоритма удаления избыточных символов:

1. Посимвольно считываем файл с исходным кодом программы;

2. Встреча пробела или знака табуляции является своего рода встречей символа-сепаратора;

3. В отличие от других символов-сепараторов не записываем в очередь лексем эти символы, т.е. игнорируем.

## 3.4 Перечень ключевых слов

Лексический анализатор преобразует исходный текст, заменяя лексические единицы лексемами для создания промежуточного представления исходной программы. Соответствие токенов и лексем приведено в таблице 3.2.

Таблица 3.2 Соответствие токенов и сепараторов с лексемами

|  |  |  |
| --- | --- | --- |
| Токен | Лексема | Пояснение |
| number, string | t | Названия типов данных языка. |
| Идентификатор | i | Длина идентификатора – 8 символов. |
| Литерал | l | Литерал любого доступного типа. |
| Шестнадцатеричный литерал | q | Литерал в шестнадцатеричном представлении. |
| function | f | Объявление функции. |
| procedure | p | Ключевое слово для скриптов – функций, не возвращающих значения. Указывается перед словом function. |
| return | e | Выход из функции/скрипта. |
| main | m | Главная функция. |
| new | n | Объявление переменной. |
| print | @ | Ввод данных. |
| criterion: | ? | Указывает начало цикла/условного оператора. |
| istrue | r | Истинная ветвь условного оператора. |
| isfalse | w | Ложная ветвь условного оператора. |
| cycle | c | Указывает на начало тела цикла. |
| newline | ^ | Оператор вывода символа перевода строки. |
| # | # | Разделение конструкций в цикле/условном операторе. |
| ; | ; | Разделение выражений. |
| , | , | Разделение параметров функций. |
| [ | [ | Начало блока/тела функции. |
| ] | ] | Закрытие блока/тела функции. |
| }  { | }  { | Знаки сдвиговых операций. |
| ( | ( | Передача параметров в функцию, приоритет операций. |
| ) | ) | Закрытие блока для передачи параметров, приоритет операций. |
| = | = | Знак присваивания. |
| +  -  \*  /  % | +  -  \*  /  % | Знаки операций. |

Продолжение таблицы 3.2

|  |  |  |
| --- | --- | --- |
| >  <  ! | >  <  ! | Знаки логических операторов |

Пример реализации таблицы лексем представлен в приложении Б.

Каждому выражению соответствует детерминированный конечный автомат, по которому происходит разбор данного выражения. На каждый автомат в массиве подаётся токен и с помощью регулярного выражения, соответствующего данному графу переходов, происходит разбор. В случае успешного разбора выражения оно записывается в таблицу лексем. Если выражение является идентификатором или литералом, информация также заносится в таблицу идентификаторов. Структура конечного автомата и пример графа перехода конечного автомата изображены на рисунках 3.3 и 3.4 соответственно.

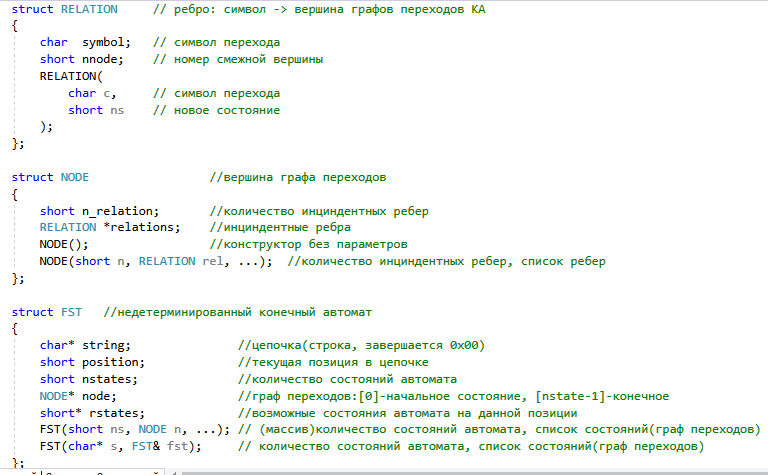


Рисунок 3.3 Структура конечного автомата

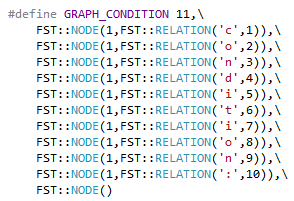


Рисунок 3.4 Пример реализации графа конечного автомата для токена criterion

## 3.5 Основные структуры данных

Основными структурами данных лексического анализатора являются таблица лексем и таблица идентификаторов. Таблица лексем содержит номер лексемы, лексему (lexema), полученную при разборе, номер строки в исходном коде (sn), и номер в таблице идентификаторов, если лексема является идентификаторов (idxTI). Таблица идентификаторов содержит имя идентификатора (id), номер в таблице лексем (idxfirstLE), тип данных (iddatatype), тип идентификатора (idtype) и значение (или параметры функций) (value). Код C++ со структурой таблицы лексем представлен на рисунке 3.3. Код C++ со структурой таблицы идентификаторов представлен на рисунке 3.4.

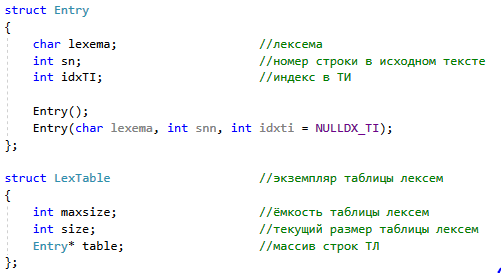


Рисунок 3.3 Структура таблицы лексем

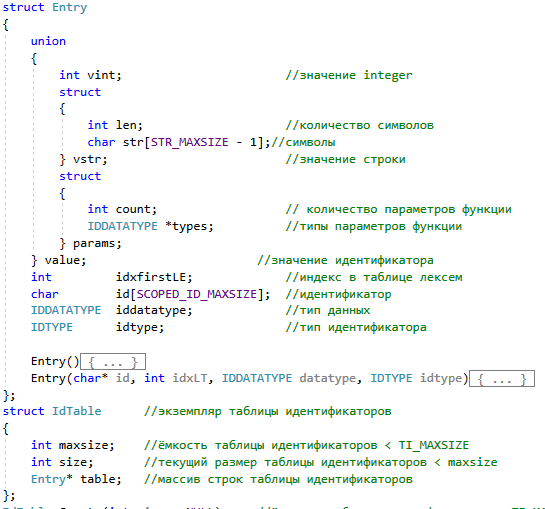


Рисунок 3.4 Структура таблицы идентификаторов

## 3.6 Принцип обработки ошибок

Для обработки ошибок лексический анализатор использует таблицу с сообщениями. Структура сообщений содержит информацию о номере сообщения, номер строки и позицию, где было вызвано сообщение в исходном коде, информацию об ошибке. При возникновении сообщения, лексический анализатор игнорирует найденную ошибку и продолжает работу с исходным кодом. Перечень сообщений представлен на рисунке 3.5.



Рисунок 3.5 Сообщения лексического анализатора

## 3.7 Структура и перечень сообщений лексического анализатора

Ошибки, возникающие в процессе трансляции программы, фиксируются в протокол, заданный входным параметрами. В случае возникновения ошибок происходит их протоколирование с номером ошибки и диагностическим сообщением. Если в процессе анализа находятся более трёх ошибок, то анализ останавливается.

## 3.8 Параметры лексического анализатора

Результаты работы лексического анализатора, а именно таблицы лексем и идентификаторов выводятся как в файл журнала, так и в командную строку.

## 3.9 Алгоритм лексического анализа

* проверяет входной поток символов программы на исходном языке на допустимость, удаляет лишние пробелы и добавляет сепаратор для вычисления номера строки для каждой лексемы;
* для выделенной части входного потока выполняется функция распознавания лексемы;
* при успешном распознавании информация о выделенной лексеме заносится в таблицу лексем и таблицу идентификаторов, и алгоритм возвращается к первому этапу;
* формирует протокол работы;
* при неуспешном распознавании выдается сообщение об ошибке.

Распознавание цепочек основывается на работе конечных автоматов. Работу конечного автомата можно проиллюстрировать с помощью графа переходов. Пример графа для цепочки «**string**» представлен на рисунке 3.2, где S0 – начальное, а S6 – конечное состояние автомата.

s t r i n g

Рисунок 3.2 Пример графа переходов для цепочки string

## 3.10 Контрольный пример

Результат работы лексического анализатора в виде таблиц лексем и идентификаторов, соответствующих контрольному примеру, представлен в приложении Б.

# 4. Разработка синтаксического анализатора

## 4.1 Структура синтаксического анализатора

Синтаксический анализатор: часть компилятора, выполняющая синтаксический анализ, то есть исходный код проверяется на соответствие правилам грамматики. Входной информацией для синтаксического анализа является таблица лексем и таблица идентификаторов. Выходной информацией является дерево разбора.

Описание структуры синтаксического анализатора языка представлено на рисунке 4.1.

Рисунок 4.1 Структура синтаксического анализатора.

## 4.2 Контекстно-свободная грамматика, описывающая синтаксис языка

В синтаксическом анализаторе транслятора языка VAD-2021 используется контекстно-свободная грамматика , где

T – множество терминальных символов (было описано в разделе 1.2 данной пояснительной записки),

N – множество нетерминальных символов (первый столбец таблицы 4.1),

P – множество правил языка (второй столбец таблицы 4.1),

S – начальный символ грамматики, являющийся нетерминалом.

Эта грамматика имеет нормальную форму Грейбах, т.к. она не леворекурсивная (не содержит леворекурсивных правил) и правила  имеют вид:

1. , где ; (или , или );
2. , где — начальный символ, при этом если такое правило существует, то нетерминал  не встречается в правой части правил.

Описание нетерминальных символов содержится в таблице 4.1.

Таблица 4.1 Таблица правил переходов нетерминальных символов

|  |  |  |
| --- | --- | --- |
| Символ | Правила | Какие правила порождает |
| S | S->tfiPTS  S->pfiPGS  S->m[K] | Стартовые правила, описывающее общую структуру программы |
| P | P->(E)  P->() | Правила для параметров объявляемых функций |
| T | T->[eV;]  T->[KeV;] | Правила для тела функций |
| G | G->[e;]  G->[Ke;] | Правила для тела скриптов |
| E | E->ti,E  E->ti | Правила для списка параметров функции |
| F | F->(N)  F->() | Правила для вызова функций (в т.ч. и в выражениях) |
| N | N->i  N>l  N->l,N  N->I,N | Правила для параметров вызываемых функций |
| R | R->rY#  R>wY#  R>cY#  R->rYwY#  R->wYrY# | Правила составления цикла/условного оператора |
| Z | Z->iLi  Z->iLl  Z->lli | Правила для условия цикла/условного оператора |
| L | L-><  L->>  L->! | Правила для логических операторов |
| A | A->+  A->-  A->\*  A->/  A->{  A->} | Правила для арифметических и сдыиговых операторов |

Продолжение таблицы 4.1

|  |  |  |
| --- | --- | --- |
| Символ | Правила | Какие правила порождает |
| V | V->l  V->i  V->q | Правила для простых выражений |
| Y | Y->[X] | Правила для тела цикла/условного выражения |
| W | W->l  W->i  W->(W)  W->(W)AW  W->iF  W->iAW  W->lAW  W->iFAW | Правила для сложных выражений |
| K | K->nti=V;K  K->nti;K  K->i=W;K  K->oV;K  K->^;K  K->&Z#RK  K->iF;K  K->nti=V;  K->nti;  K->i=W;  K->oV;  K->^;  K->&Z#R  K->iF; | Программные конструкции |
| X | X->i=W;X  X->oV;X  X->^;X  X->iF;X  X->i=W;  X->oV;  X->^;  X->iF; | Программные конструкции внутри цикла/условного оператора |

## 4.3 Построение конечного магазинного автомата

Конечный автомат с магазинной памятью представляет собой семерку. Подробное описание компонентов магазинного автомата представлено в таблице 4.2.

Таблица 4.2 – Описание компонентов магазинного автомата

|  |  |  |
| --- | --- | --- |
| Компонента | Определение | Описание |
|  | Множество состояний автомата | Состояние автомата представляет из себя структуру, содержащую позицию на входной ленте, номера текущего правила и цепочки и стек автомата |
|  | Алфавит входных символов | Алфавит представляет из себя множества терминальных и нетерминальных символов, описание которых содержится в таблица 3.1 и 4.1. |
|  | Алфавит специальных магазинных символов | Алфавит магазинных символов содержит стартовый символ и маркер дна стека (представляет из себя символ $) |
|  | Функция переходов автомата | Функция представляет из себя множество правил грамматики, описанных в таблице 4.1. |
|  | Начальное состояние автомата | Состояние, которое приобретает автомат в начале своей работы. Представляется в виде стартового правила грамматики |
|  | Начальное состояние магазина автомата | Символ маркера дна стека $ |
|  | Множество конечных состояний | Конечные состояние заставляют автомат прекратить свою работу. Конечным состоянием является пустой магазин автомата и совпадение позиции на входной ленте автомата с размером ленты |

## 4.4 Основные структуры данных

Основные структуры данных синтаксического анализатора представляются в виде структуры магазинного конечного автомата, выполняющего разбор исходной ленты, и структуры грамматики Грейбах, описывающей синтаксические правила языка VAD-2021 . Данные структуры в приложении В.

## 4.5 Описание алгоритма синтаксического разбора

Принцип работы автомата следующий:

1. В магазин записывается стартовый символ;
2. На основе полученных ранее таблиц формируется входная лента;
3. Запускается автомат;
4. Выбирается цепочка, соответствующая нетерминальному символу, записывается в магазин в обратном порядке;
5. Если терминалы в стеке и в ленте совпадают, то данный терминал удаляется из ленты и стека. Иначе возвращаемся в предыдущее сохраненное состояние и выбираем другую цепочку нетерминала;
6. Если в магазине встретился нетерминал, переходим к пункту 4;
7. Если наш символ достиг дна стека, и лента в этот момент пуста, то синтаксический анализ выполнен успешно. Иначе генерируется исключение.

## 4.6 Структура и перечень сообщений синтаксического анализатора

Перечень сообщений синтаксического анализатора представлен на рисунке 4.3.

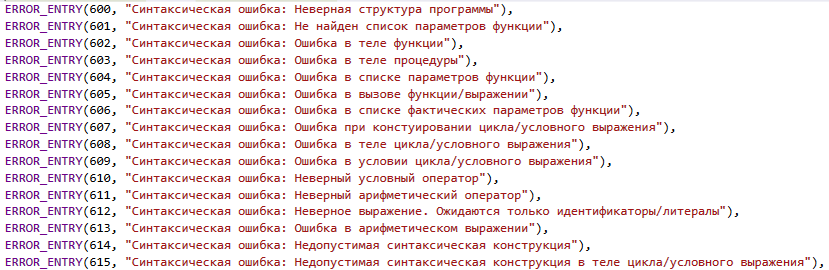


Рисунок 4.3 Сообщения синтаксического анализатора

## 4.7. Параметры синтаксического анализатора и режимы его работы

Входной информацией для синтаксического анализатора является таблица лексем и идентификаторов. Кроме того используется описание грамматики в форме Грейбах. Результаты работы лексического разбора, а именно дерево разбора и протокол работы автомата с магазинной памятью выводятся в журнал работы программы.

## 4.8. Принцип обработки ошибок

Синтаксический анализатор выполняет разбор исходной последовательности лексем до тех пор, пока не дойдёт до конца цепочки лексем или не найдёт ошибку. Тогда анализ останавливается и выводится сообщение об ошибке (если она найдена). Если в процессе анализа находятся более трёх ошибок, то анализ останавливается.

## 4.9. Контрольный пример

Результаты работы лексического разбора, а именно дерево разбора и протокол работы автомата с магазинной памятью приведены в приложении В.

# 5 Разработка семантического анализатора

## 5.1 Структура семантического анализатора

Семантический анализатор принимает на свой вход результаты работ лексического и синтаксического анализаторов, то есть таблицы лексем, идентификаторов и результат работы синтаксического анализатора, то есть дерево разбора, и последовательно ищет необходимые ошибки. Некоторые проверки (такие как проверка на единственность точки входа, проверка на предварительное объявление переменной) осуществляются в процессе лексического анализа. Общая структура обособленно работающего (не параллельно с лексическим анализом) семантического анализатора представлена на рисунке 5.1.



Рисунок 5.1. Структура семантического анализатора

## 5.2 Функции семантического анализатора

Семантический анализатор выполняет проверку на основные правила языка (семантики языка), которые описаны в разделе 1.16.

## 5.3 Структура и перечень сообщений семантического анализатора

Сообщения, формируемые семантическим анализатором, представлены на рисунке 5.1.

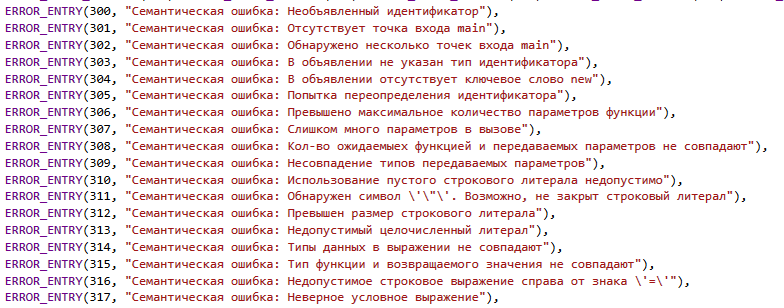


Рисунок 5.2 Перечень сообщений семантического анализатора

## 5.4 Принцип обработки ошибок

Ошибки, возникающие в процессе трансляции программы, фиксируются в протокол, заданный входным параметрами. В случае возникновения ошибок происходит их протоколирование с номером ошибки и диагностическим сообщением. Анализ останавливается после того, как будут найдены все ошибки.

## 5.5 Контрольный пример

Соответствие примеров некоторых ошибок в исходном коде и диагностических сообщений об ошибках приведено в таблице 5.1.

Таблица 5.1. Примеры диагностики ошибок

|  |  |
| --- | --- |
| Исходный код | Текст сообщения |
| main  [  number x = 9;  print x;  ] | Ошибка №304: Семантическая ошибка: В объявлении отсутствует ключевое слово new Строка: 2 |
| main  [  new number x = 9;  new string y =x;  ] | Ошибка №314: Семантическая ошибка: Типы данных в выражении не совпадают Строка: 3 |
| main  [  new number x = 9;  ]  main  [  new string y = "qwerty";  ] | Ошибка №302: Семантическая ошибка: Обнаружено несколько точек входа main Строка: 0 |

# 6. Вычисление выражений

## 6.1 Выражения, допускаемые языком

В языке VAD-2021 допускаются вычисления выражений целочисленного типа данных с поддержкой вызова функций внутри выражений. Приоритет операций представлен на таблице 6.1.

Таблица 6.1. Приоритеты операций

|  |  |
| --- | --- |
| Операция | Значение приоритета |
| ( ) | 3 |
| \* | 2 |
| / | 2 |
| % | 2 |
| + | 1 |
| - | 1 |
| } | 0 |
| { | 0 |

## 6.2 Польская запись и принцип её построения

Все выражения языка VAD-2021 преобразовываются к обратной польской записи.

Польская запись – это альтернативный способ записи арифметических выражений, преимущество которого состоит в отсутствии скобок. Существует два типа польской записи: префиксная и постфиксная. Отличие их от классического, инфиксного способа заключается в том, что знаки операций пишутся не между, а, соответственно, до или после аргументов. Алгоритм построения польской записи:

* исходная строка: выражение;
* результирующая строка: польская запись;
* стек: пустой;
* исходная строка просматривается слева направо;
* операнды переносятся в результирующую строку;
* операция записывается в стек, если стек пуст;
* операция выталкивает все операции с большим или равным приоритетом в результирующую строку;
* отрывающая скобка помещается в стек;
* закрывающая скобка выталкивает все операции до открывающей скобки, после чего обе скобки уничтожаются.

## 6.3 Программная реализация обработки выражений

Программная реализация алгоритма преобразования выражений к польской записи представлена в приложении Г.

## 6.4 Контрольный пример

Пример преобразования выражений из контрольных примеров к обратной польской записи представлен в таблице 6.2. Преобразование выражений в формат польской записи необходимо для построения более простых алгоритмов их вычисления и преобразования к ассемблерному коду. В приложении Г приведены изменённые таблицы лексем и идентификаторов, отображающие результаты преобразования выражений в польский формат.

Таблица 6.2. Преобразование выражений к ПОЛИЗ

|  |  |
| --- | --- |
| Выражение | Обратная польская запись для выражения |
| i[2]=(((l[3]+l[4])-i[0])\*l[5])/l[6]; | i[2]=l[3]l[4]+i[0]-l[5]\*l[6]/ |
| i[23]=(i[23]+l[26])\*l[26] | i[23]=i[23]l[26]+l[26]\* |
| i[3]=(((l[4]+l[5])-i[0])\*l[6]) | i[3]=l[4]l[5]+i[0]-l[6]\* |

# 7. Генерация кода

## 7.1 Структура генератора кода

В языке VAD-2021 генерация кода является заключительным этапом трансляции. Генератор принимает на вход таблицы лексем и идентификаторов, полученные в результате лексического анализа. В соответствии с таблицей лексем строится выходной файл на языке ассемблера, который будет являться результатом работы транслятора. В случае возникновения ошибок генерация кода не будет осуществляться. Структура генератора кода VAD-2021 представлена на рисунке 7.1.



Рисунок 7.1 – Структура генератора кода

## 7.2 Представление типов данных в оперативной памяти

Элементы таблицы идентификаторов расположены сегментах .data и .const языка ассемблера. Соответствия между типами данных идентификаторов на языке VAD-2021 и на языке ассемблера приведены в таблице 7.1.

Таблица 7.1 – Соответствия типов идентификаторов языка VAD-2021 и языка ассемблера

|  |  |  |
| --- | --- | --- |
| Тип идентификатора на языке VAD-2021 | Тип идентификатора на языке ассемблера | Пояснение |
| number | sdword | Хранит целочисленный тип данных. |
| string | dword | Хранит указатель на начало строки. Строка должна завешаться нулевым символом. |

## 7.3 Статическая библиотека

В языке VAD-2021 предусмотрена статическая библиотека. Статическая библиотека содержит функции, написанные на языке C++. Объявление функций статической библиотеки генерируется автоматически в коде ассемблера. Объявление функций статической библиотеки генерируется автоматически.

Таблица 7.3 – Функции статической библиотеки

|  |  |
| --- | --- |
| Функция | Назначение |
| void outlstr(char\* str) | Вывод на консоль строки str |
| void outnum(int num) | Вывод на консоль целочисленной переменной num |
| int length(char\* buffer, char\* str) | Вычисление длины строки |
| char\* strcopy(char\* str) | Копирование строки str |
| int atoii(char\* ptr) | Преобразование строки в число |

## 7.4 Особенности алгоритма генерации кода

В языке VAD-2021 генерация кода строится на основе таблиц лексем и идентификаторов. Общая схема работы генератора кода представлена на рисунке 7.2.

Рисунок 7.2 Структура генератора кода

## 7.5 Входные параметры генератора кода

На вход генератору кода поступают таблицы лексем и идентификаторов исходного код программы на языке VAD-2021. Результаты работы генератора кода выводятся в файл с расширением .asm.

## 7.6 Контрольный пример

Результат генерации ассемблерного кода на основе контрольного примера из приложения А приведен в приложении Д. Результат работы контрольного примера приведён на рисунке 7.2.

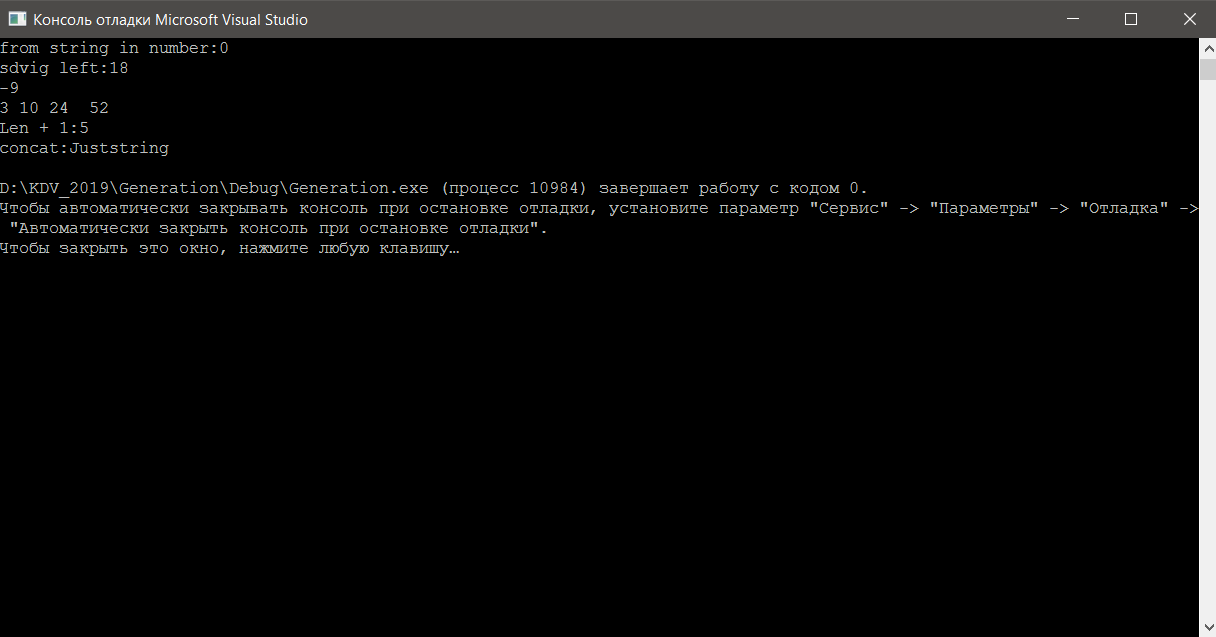


Рисунок 7.2 Результат работы программы на языке VAD-2021

# 8. Тестирование транслятора

## 8.1 Тестирование проверки на допустимость символов

В языке VAD-2021 не разрешается использовать запрещённые входным алфавитом символы. Результат использования запрещённого символа показан в таблице 8.1.

Таблица 8.1 - Тестирование проверки на допустимость символов

|  |  |
| --- | --- |
| Исходный код | Диагностическое сообщение |
| main [ё] | Ошибка №200: Лексическая ошибка: Недопустимый символ в исходном файле(-in) Строка: 2 Позиция в строке: 2 |

## 8.2 Тестирование лексического анализатора

На этапе лексического анализа в языке VAD-2021 могут возникнуть ошибки, описанные в пункте 3.7. Результаты тестирования лексического анализатора показаны в таблице 8.2.

Таблица 8.2 - Тестирование лексического анализатора

|  |  |
| --- | --- |
| Исходный код | Диагностическое сообщение |
| main  [new number w11;] | Ошибка №201: Лексическая ошибка: Неизвестная последовательность символов Строка: 3 |

## 8.3 Тестирование синтаксического анализатора

На этапе синтаксического анализа в языке VAD-2021 могут возникнуть ошибки, описанные в пункте 4.6. Результаты тестирования синтаксического анализатора показаны в таблице 8.3.

Таблица 8.3 - Тестирование синтаксического анализатора

|  |  |
| --- | --- |
| Исходный код | Диагностическое сообщение |
| main new number x; ] | Ошибка №600: строка 1, Синтаксическая ошибка: Неверная структура программы |
| string function fi([]  main[] | Ошибка №601: строка 1, Синтаксическая ошибка: Не найден список параметров функции |
| string function fi()  [newline; print]  main[] | Ошибка №602: строка 1, Синтаксическая ошибка: Ошибка в теле функции |
| procedure function fi()  [newline print;]  main[] | Ошибка №603: строка 1, Синтаксическая ошибка: Ошибка в теле скрипта |

Продолжение таблицы 8.3

|  |  |
| --- | --- |
| Исходный код | Диагностическое сообщение |
| procedure function fi(number number)[]  main[] | Ошибка №604: строка 1, Синтаксическая ошибка: Ошибка в списке параметров функции |
| string function fi(number x)[return 3;]  main [newline;fi(5,5; newline;] | Ошибка №605: строка 2, Синтаксическая ошибка: Ошибка в вызове функции/выражении |
| string function fi(number x)[return 3;]  main [newline;fi(5,5,5 5);] | Ошибка №606: строка 2, Синтаксическая ошибка: Ошибка в списке фактических параметров функции |
| main [new number x; criterion: x > 2 # cycle #] | Ошибка №607: строка 1, Синтаксическая ошибка: Ошибка при конструировании цикла/условного выражения |
| main [new number x; criterion: x > 2 # cycle #] | Ошибка №608: строка 1, Синтаксическая ошибка: Ошибка в теле цикла/условного выражения |
| main [criterion: 1 = 2 #] | Ошибка №609: строка 1, Синтаксическая ошибка: Ошибка в условии цикла/условного выражения |
| main [criterion: 1 = 2 #] | Ошибка №610: строка 1, Синтаксическая ошибка: Неверный условный оператор |
| main [new number x; x = x ! x;] | Ошибка №611: строка 1, Синтаксическая ошибка: Неверный арифметический оператор |
| main [new number x; print new;] | Ошибка №612: строка 1, Синтаксическая ошибка: Неверное выражение. Ожидаются только идентификаторы/литералы |
| main [new number x; x = 1 ++;] | Ошибка №613: строка 1, Синтаксическая ошибка: Ошибка в арифметическом выражении |
| main [newline; 4;] | Ошибка №614: строка 1, Синтаксическая ошибка: Недопустимая синтаксическая конструкция |
| main [new number a; criterion: a < 3 # istrue [newline; 3;] #] | Ошибка №615: строка 1, Синтаксическая ошибка: Недопустимая синтаксическая конструкция в теле цикла/условного выражения |

## 8.4 Тестирование семантического анализатора

Семантический анализ в языке VAD-2021 содержит множество проверок по семантическим правилам, описанным в пункте 1.16. Итоги тестирования семантического анализатора на корректное обнаружение семантических ошибок приведены в таблице 8.4.

Таблица 8.4 - Тестирование семантического анализатора

|  |  |
| --- | --- |
| Исходный код | Диагностическое сообщение |
| main [ a = 1 ] | Ошибка №300: Семантическая ошибка: Необъявленный идентификатор. Строка: 1 |
| string function fi()[] | Ошибка №301: Семантическая ошибка: Отсутствует точка входа main. Строка: 0 |
| main[]  main[] | Ошибка №302: Семантическая ошибка: Обнаружено несколько точек входа main. Строка: 0 |
| main[a = 1;] | Ошибка №304: Семантическая ошибка: В объявлении отсутствует ключевое слово new. Строка: 1 |
| main[new number t;  new string t;] | Ошибка №305: Семантическая ошибка: Попытка переопределения идентификатора. Строка: 3 |
| procedure function fi()[]  main[fi(“a”,”b”,”c”,”d”] | Ошибка №307: Семантическая ошибка: Слишком много параметров в вызове. Строка: 1 |
| string function fi(string x, string y, string z, string s)  main[] | Ошибка №306: Семантическая ошибка: Превышено максимальное количество параметров функции. Строка: 1 |
| string function fi(string x)[return "a";]  main[fi("a", "b");] | Ошибка №308: Семантическая ошибка: Кол-во ожидаемых функцией и передаваемых параметров не совпадают. Строка: 2 |
| string function fi(string x)[return "a";]  main[fi("a", "b");] | Ошибка №309: Семантическая ошибка: Несовпадение типов передаваемых параметров Строка: 2 |
| main[new string x="";] | Ошибка №310: Семантическая ошибка: Использование пустого строкового литерала недопустимо. Строка: 1 |
| main[new string x=";] | Ошибка №311: Семантическая ошибка: Обнаружен символ '"'. Строка: 1 |
| main[new number x=99999999999999999;] | Ошибка №313: Семантическая ошибка: Недопустимый целочисленный литерал. Строка: 1 |
| main[new number x; x = 5 + "abc";] | Ошибка №314: Семантическая ошибка: Типы данных в выражении не совпадают. Строка: 1 |
| string function fi()[return 5;]  main[newline;] | Ошибка №315: Семантическая ошибка: Тип функции и возвращаемого значения не совпадают. Строка: 1 |
| main[new string x; x = "abc" + "d";] | Ошибка №316: Семантическая ошибка: Недопустимое строковое выражение справа от знака '='. Строка: 1 |
| main  [criterion: "string"& 6#  istrue[print "string";]] | Ошибка №317: Семантическая ошибка: Неверное условное выражение. Строка: 1 |
| main[new number a =5;  a = a/0; print a;] | Ошибка №318: Семантическая ошибка: Деление на ноль. Строка: 4 |

# Заключение

В ходе выполнения курсовой работы был разработан транслятор и генератор кода для языка программирования VAD-2021 со всеми необходимыми компонентами. Таким образом, были выполнены основные задачи данной курсовой работы:

1. Сформулирована спецификация языка VAD-2021;
2. Разработаны конечные автоматы и важные алгоритмы на их основе для эффективной работы лексического анализатора;
3. Осуществлена программная реализация лексического анализатора, распознающего допустимые цепочки спроектированного языка;
4. Разработана контекстно-свободная, приведённая к нормальной форме Грейбах, грамматика для описания синтаксически верных конструкций языка;
5. Осуществлена программная реализация синтаксического анализатора;
6. Разработан семантический анализатор, осуществляющий проверку используемых инструкций на соответствие логическим правилам;
7. Разработан транслятор кода на язык ассемблера;
8. Проведено тестирование всех вышеперечисленных компонентов.

Окончательная версия языка VAD-2021 включает:

1. 2 типа данных;
2. Поддержка операторов вывода и перевода строки;
3. Возможность вызова функций стандартной библиотеки;
4. Наличие 5 арифметических операторов для вычисления выражений;
5. Поддержка функций, скриптов, операторов цикла и условия;
6. Структурированная и классифицированная система для обработки ошибок пользователя.

Проделанная работа позволила получить необходимое представление о структурах и процессах, использующихся при построении трансляторов, а также основные различия и преимущества тех или иных средств трансляции.

# Список использованных источников

1. Курс лекций по ЯП / Наркевич А.С.

2. Ахо, А. Компиляторы: принципы, технологии и инструменты / А. Ахо, Р. Сети, Дж. Ульман. – M.: Вильямс, 2003. – 768с.

3. Герберт, Ш. Справочник программиста по C/C++ / Шилдт Герберт. – 3-е изд. – Москва : Вильямс, 2003. - 429 с.

4. Прата, С. Язык программирования С++. Лекции и упражнения / С. Прата. – М., 2006 — 1104 c.

5. Креншоу, Д. Давайте создадим компилятор! / Д. Креншоу. – 3-е изд. – Моска: 1995 – 73 с.

6. Пратт, Т. Языки программирования: разработка и реализация / Т. Пратт – 4-е изд., 2002 – 690 с.

# Приложение А

Листинг 1 – Исходный код программы на языка VAD-2021

number function min(number x, number y)

[

new number res;

criterion: x < y #

istrue [res = x;]

isfalse [res = y;]#

return res;

]

procedure function stand (string a, string b)

[

new number k;

k = length(a)+1;

print "Len + 1:";

print k;

newline;

new string str;

str = strcopy(a);

print "strcopy:";

print str;

newline;

return;

]

main

[

new number x = 9;

new number y = -9;

new string strx = "Just";

new string stry = "string";

new string strz = "70";

new number e;

print "from string in number:";

e = atoii(strz);

print e;

newline;

new number result;

result = x{1;

print "sdvig left:";

print result;

newline;

new number t;

t = min (x,y);

print t;

newline;

new number ab = 3;

new number d = q120;

criterion: ab ! 52 #

cycle [

print ab;

print " ";

ab = (ab + 2)\*2;

]#

print " ";

print ab;

newline;

stand(strx,stry);

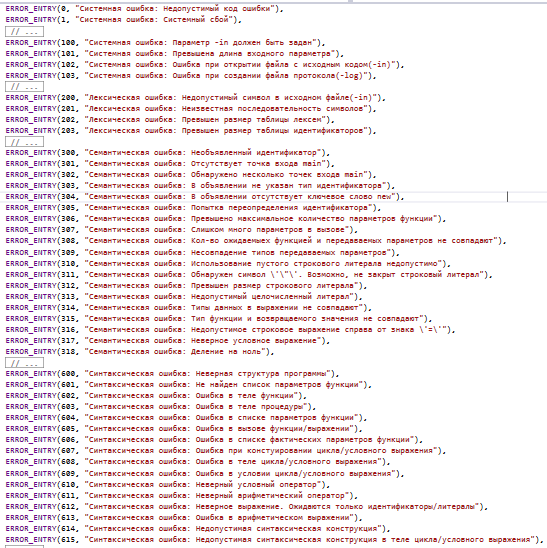
]

Рисунок 1 Таблица ошибок языка VAD-2021

# Приложение Б

Листинг 1 Таблица идентификаторов контрольного примера

| N |СТРОКА В ТЛ| ТИП ИДЕНТИФИКАТОРА | ИМЯ | ЗНАЧЕНИЕ (ПАРАМЕТРЫ)

| 0 | 2 | number function | min | P0:NUMBER | P1:NUMBER |

| 1 | 5 | number parameter | minx |

| 2 | 8 | number parameter | miny |

| 3 | 13 | number variable | minres |0

| 4 | 41 | proc function | stand | P0:STRING | P1:STRING |

| 5 | 44 | string parameter | standa |

| 6 | 47 | string parameter | standb |

| 7 | 52 | number variable | standk |0

| 8 | 56 | number LIB FUNC | length | P0:STRING |

| 9 | 68 | string variable | standstr |[0]

| 10 | 72 | string LIB FUNC | strcopy | P0:STRING

| 11 | 91 | number variable | mainx |0

| 12 | 93 | number literal | LTRL1 |9

| 13 | 97 | number variable | mainy |0

| 14 | 99 | number literal | LTRL2 |-9

| 15 | 103 | string variable | mainstrx |[0]

| 16 | 105 | string literal | LTRL3 |[4]Just

| 17 | 109 | string variable | mainstry |[0]

| 18 | 111 | string literal | LTRL4 |[6]string

| 19 | 115 | number variable | maint |0

| 20 | 133 | number variable | mainab |0

| 21 | 135 | number literal | LTRL5 |3

| 22 | 140 | number literal | LTRL6 |52

| 23 | 148 | string literal | LTRL7 |[1]

| 24 | 155 | number literal | LTRL8 |2

Листинг 2 Таблица лексем после контрольного примера

| N | ЛЕКСЕМА | СТРОКА | ИНДЕКС В ТИ |

| 0 | t | 1 | |

| 1 | f | 1 | |

| 2 | i | 1 | 0 |

| 3 | ( | 1 | |

| 4 | t | 1 | |

| 5 | i | 1 | 1 |

| 6 | , | 1 | |

| 7 | t | 1 | |

| 8 | i | 1 | 2 |

| 9 | ) | 1 | |

| 10 | [ | 2 | |

| 11 | n | 3 | |

| 12 | t | 3 | |

| 13 | i | 3 | 3 |

| 14 | ; | 3 | |

| 15 | ? | 4 | |

| 16 | i | 4 | 1 |

| 17 | < | 4 | |

| 18 | i | 4 | 2 |

| 19 | # | 4 | |

| 20 | w | 5 | |

| 21 | [ | 5 | |

| 22 | i | 5 | 3 |

| 23 | = | 5 | |

| 24 | i | 5 | 1 |

| 25 | ; | 5 | |

| 26 | ] | 5 | |

| 27 | r | 6 | |

| 28 | [ | 6 | |

| 29 | i | 6 | 3 |

| 30 | = | 6 | |

| 31 | i | 6 | 2 |

| 32 | ; | 6 | |

| 33 | ] | 6 | |

| 34 | # | 6 | |

| 35 | e | 7 | |

| 36 | i | 7 | 3 |

| 37 | ; | 7 | |

| 38 | ] | 8 | |

| 39 | p | 9 | |

| 40 | f | 9 | |

| 41 | i | 9 | 4 |

| 42 | ( | 9 | |

| 43 | t | 9 | |

| 44 | i | 9 | 5 |

| 45 | , | 9 | |

| 46 | t | 9 | |

| 47 | i | 9 | 6 |

| 48 | ) | 9 | |

| 49 | [ | 10 | |

| 50 | n | 11 | |

| 51 | t | 11 | |

| 52 | i | 11 | 7 |

| 53 | ; | 11 | |

| 54 | i | 12 | 7 |

| 55 | = | 12 | |

| 56 | i | 12 | 8 |

| 57 | ( | 12 | |

| 58 | i | 12 | 5 |

| 59 | ) | 12 | |

| 60 | ; | 12 | |

| 61 | o | 13 | |

| 62 | i | 13 | 7 |

| 63 | ; | 13 | |

| 64 | ^ | 14 | |

| 65 | ; | 14 | |

| 66 | n | 15 | |

| 67 | t | 15 | |

| 68 | i | 15 | 9 |

| 69 | ; | 15 | |

| 70 | i | 16 | 9 |

| 71 | = | 16 | |

| 72 | i | 16 | 10 |

| 73 | ( | 16 | |

| 74 | i | 16 | 5 |

| 75 | , | 16 | |

| 76 | i | 16 | 6 |

| 77 | ) | 16 | |

| 78 | ; | 16 | |

| 79 | o | 17 | |

| 80 | i | 17 | 9 |

| 81 | ; | 17 | |

| 82 | ^ | 18 | |

| 83 | ; | 18 | |

| 84 | e | 19 | |

| 85 | ; | 19 | |

| 86 | ] | 20 | |

| 87 | m | 21 | |

| 88 | [ | 22 | |

| 89 | n | 23 | |

| 90 | t | 23 | |

| 91 | i | 23 | 11 |

| 92 | = | 23 | |

| 93 | l | 23 | 12 |

| 94 | ; | 23 | |

| 95 | n | 24 | |

| 96 | t | 24 | |

| 97 | i | 24 | 13 |

| 98 | = | 24 | |

| 99 | l | 24 | 14 |

|100 | ; | 24 | |

|101 | n | 25 | |

|102 | t | 25 | |

|103 | i | 25 | 15 |

|104 | = | 25 | |

|105 | l | 25 | 16 |

|106 | ; | 25 | |

|107 | n | 26 | |

|108 | t | 26 | |

|109 | i | 26 | 17 |

|110 | = | 26 | |

|111 | l | 26 | 18 |

|112 | ; | 26 | |

|113 | n | 27 | |

|114 | t | 27 | |

|115 | i | 27 | 19 |

|116 | ; | 27 | |

|117 | i | 28 | 19 |

|118 | = | 28 | |

|119 | i | 28 | 0 |

|120 | ( | 28 | |

|121 | i | 28 | 11 |

|122 | , | 28 | |

|123 | i | 28 | 13 |

|124 | ) | 28 | |

|125 | ; | 28 | |

|126 | o | 29 | |

|127 | i | 29 | 19 |

|128 | ; | 29 | |

|129 | ^ | 30 | |

|130 | ; | 30 | |

|131 | n | 31 | |

|132 | t | 31 | |

|133 | i | 31 | 20 |

|134 | = | 31 | |

|135 | l | 31 | 21 |

|136 | ; | 31 | |

|137 | ? | 32 | |

|138 | i | 32 | 20 |

|139 | ! | 32 | |

|140 | l | 32 | 22 |

|141 | # | 32 | |

|142 | c | 33 | |

|143 | [ | 33 | |

|144 | o | 34 | |

|145 | i | 34 | 20 |

|146 | ; | 34 | |

|147 | o | 35 | |

|148 | l | 35 | 23 |

|149 | ; | 35 | |

|150 | i | 36 | 20 |

|151 | = | 36 | |

|152 | ( | 36 | |

|153 | i | 36 | 20 |

|154 | + | 36 | |

|155 | l | 36 | 24 |

|156 | ) | 36 | |

|157 | \* | 36 | |

|158 | l | 36 | 24 |

|159 | ; | 36 | |

|160 | ] | 37 | |

|161 | # | 37 | |

|162 | o | 38 | |

|163 | l | 38 | 23 |

|164 | ; | 38 | |

|165 | o | 39 | |

|166 | i | 39 | 20 |

|167 | ; | 39 | |

|168 | ^ | 40 | |

|169 | ; | 40 | |

|170 | i | 41 | 4 |

|171 | ( | 41 | |

|172 | i | 41 | 15 |

|173 | , | 41 | |

|174 | i | 41 | 17 |

|175 | ) | 41 | |

|176 | ; | 41 | |

|177 | ] | 42 | |

# Приложение В

Листинг 1 Грамматика языка VAD-2021

reibach greibach(NS('S'), TS('$'), 16,

Rule(NS('S'), GRB\_ERROR\_SERIES, 3, // Неверная структура программы

Rule::Chain(6, TS('t'), TS('f'), TS('i'), NS('P'), NS('T'), NS('S')),

Rule::Chain(6, TS('p'), TS('f'), TS('i'), NS('P'), NS('G'), NS('S')),

Rule::Chain(4, TS('m'), TS('['), NS('K'), TS(']'))

),

Rule(NS('T'), GRB\_ERROR\_SERIES + 2, 2, // Ошибка в теле функции

Rule::Chain(5, TS('['), TS('e'), NS('V'), TS(';'), TS(']')),

Rule::Chain(6, TS('['), NS('K'), TS('e'), NS('V'), TS(';'), TS(']'))

),

Rule(NS('G'), GRB\_ERROR\_SERIES + 3, 2, // Ошибка в теле скритпа

Rule::Chain(4, TS('['), TS('e'), TS(';'), TS(']')),

Rule::Chain(5, TS('['), NS('K'), TS('e'), TS(';'), TS(']'))

Rule(NS('P'), GRB\_ERROR\_SERIES + 1, 2, // Не найден список параметров функции

Rule::Chain(3, TS('('), NS('E'), TS(')')),

Rule::Chain(2, TS('('), TS(')'))

),

Rule(NS('E'), GRB\_ERROR\_SERIES + 4, 2, // Ошибка в списке параметров функции

Rule::Chain(4, TS('t'), TS('i'), TS(','), NS('E')),

Rule::Chain(2, TS('t'), TS('i'))

),

Rule(NS('F'), GRB\_ERROR\_SERIES + 5, 2, // Ошибка в вызове функции

Rule::Chain(3, TS('('), NS('N'), TS(')')),

Rule::Chain(2, TS('('), TS(')'))

),

Rule(NS('N'), GRB\_ERROR\_SERIES + 6, 4, // Ошибка в списке параметров функции

Rule::Chain(1, TS('i')),

Rule::Chain(1, TS('l')),

Rule::Chain(3, TS('i'), TS(','), NS('N')),

Rule::Chain(3, TS('l'), TS(','), NS('N'))

),

Rule(NS('R'), GRB\_ERROR\_SERIES + 7, 5, // Ошибка при констуировании цикла/условного выражения

Rule::Chain(3, TS('r'), NS('Y'), TS('#')),

Rule::Chain(3, TS('w'), NS('Y'), TS('#')),

Rule::Chain(3, TS('c'), NS('Y'), TS('#')),

Rule::Chain(5, TS('r'), NS('Y'), TS('w'), NS('Y'), TS('#')),

Rule::Chain(5, TS('w'), NS('Y'), TS('r'), NS('Y'), TS('#'))

),

Rule(NS('Y'), GRB\_ERROR\_SERIES + 8, 1, // Ошибка в теле цикла/условного выражения

Rule::Chain(3, TS('['), NS('X'), TS(']'))

),

Rule(NS('Z'), GRB\_ERROR\_SERIES + 9, 3, // Ошибка в условии цикла/условного выражения

Rule::Chain(3, TS('i'), NS('L'), TS('i')),

Rule::Chain(3, TS('i'), NS('L'), TS('l')),

Rule::Chain(3, TS('l'), NS('L'), TS('i'))

),

Листинг 1(продолжение) Грамматика языка VAD-2021

Rule(NS('L'), GRB\_ERROR\_SERIES + 10, 4,// Неверный условный оператор

Rule::Chain(1, TS('<')),

Rule::Chain(1, TS('>')),

Rule::Chain(1, TS('&')),

Rule::Chain(1, TS('!'))

),

Rule(NS('A'), GRB\_ERROR\_SERIES + 11, 4, // Неверный арифметический оператор

Rule::Chain(1, TS('+')),

Rule::Chain(1, TS('-')),

Rule::Chain(1, TS('\*')),

Rule::Chain(1, TS('/'))

),

Rule(NS('V'), GRB\_ERROR\_SERIES + 12, 2, // Неверное выражение. Rule::Chain(1, TS('l')),

Rule::Chain(1, TS('i'))

),

Rule(NS('W'), GRB\_ERROR\_SERIES + 13, 8,// Ошибка в арифметичском выражении

Rule::Chain(1, TS('i')),

Rule::Chain(1, TS('l')),

Rule::Chain(3, TS('('), NS('W'), TS(')')),

Rule::Chain(5, TS('('), NS('W'), TS(')'), NS('A'), NS('W')),

Rule::Chain(2, TS('i'), NS('F')),

Rule::Chain(3, TS('i'), NS('A'), NS('W')),

Rule::Chain(3, TS('l'), NS('A'), NS('W')),

Rule::Chain(4, TS('i'), NS('F'), NS('A'), NS('W'))

),

Rule(NS('K'), GRB\_ERROR\_SERIES + 14, 14,// Недопустимая синтаксическая конструкция

Rule::Chain(7, TS('n'), TS('t'), TS('i'), TS('='), NS('V'), TS(';'), NS('K')), // декларация + присваивание

Rule::Chain(5, TS('n'), TS('t'), TS('i'), TS(';'), NS('K')), // декларация

Rule::Chain(5, TS('i'), TS('='), NS('W'), TS(';'), NS('K')), // присваивание

Rule::Chain(4, TS('o'), NS('V'), TS(';'), NS('K')), // вывод

Rule::Chain(3, TS('^'), TS(';'), NS('K')), // перевод строки

Rule::Chain(5, TS('?'), NS('Z'), TS('#'), NS('R'), NS('K')), // criterion

Rule::Chain(4, TS('i'), NS('F'), TS(';'), NS('K')), // вызов функции

Rule::Chain(6, TS('n'), TS('t'), TS('i'), TS('='), NS('V'), TS(';')), // декларация + присваивание

Rule::Chain(4, TS('i'), TS('='), NS('W'), TS(';')), // присваивание

Rule::Chain(4, TS('n'), TS('t'), TS('i'), TS(';')), // декларация

Rule::Chain(3, TS('o'), NS('V'), TS(';')), // вывод

Rule::Chain(2, TS('^'), TS(';')), // перевод строки

Rule::Chain(4, TS('?'), NS('Z'), TS('#'), NS('R')), // criterion

Rule::Chain(3, TS('i'), NS('F'), TS(';')) // вызов функции

),

Rule(NS('X'), GRB\_ERROR\_SERIES + 15, 8, // Недопустимая синтаксическая конструкция в теле цикла/условного выражения

Rule::Chain(5, TS('i'), TS('='), NS('W'), TS(';'), NS('X')), //присваивание

Rule::Chain(4, TS('o'), NS('V'), TS(';'), NS('X')), // вывод

Rule::Chain(3, TS('^'), TS(';'), NS('X')), // перевод строки

Rule::Chain(4, TS('i'), NS('F'), TS(';'), NS('X')), // вызов функции

Rule::Chain(4, TS('i'), TS('='), NS('W'), TS(';')), // присваивание

Rule::Chain(3, TS('o'), NS('V'), TS(';')), // вывод

Rule::Chain(2, TS('^'), TS(';')), // перевод строки

Rule::Chain(3, TS('i'), NS('F'), TS(';')) // вызов функции

)

);

Листинг 2 Структура магазинного автомата

struct Mfst { //магазинный автомат

enum RC\_STEP //шаг автомата

{

NS\_OK, //найдено правило и цепочка, цепочка записана в стек

NS\_NORULE, //не найдено правило грамматики (ошибки в грамматике)

NS\_NORULECHAIN,//не найдена подходящая цепочка правила NS\_ERROR, //неизвестный нетерминальный символ грамматики

TS\_OK, //текущий символ ленты == вершине стека, продвинулась лента TS\_NOK,//текущий символ ленты != вершине стека, восстановлено состояние

LENTA\_END, //текущая позиция ленты >= lenta\_size

SURPRISE //неожиданный код возврата ( ошибка в step)

};

struct MfstDiagnosis //диагностика

{

short lenta\_position; //позиция на ленте

RC\_STEP rc\_step; //код завершения шага

short nrule; //номер правила

short nrule\_chain; //номер цепочки правила

MfstDiagnosis();

MfstDiagnosis( //диагностика

short plenta\_position, //позиция на ленте

RC\_STEP prc\_step, //код завершения шага

short pnrule, //номер правила

short pnrule\_chain //номер цепочки правила

);

}

diagnosis[MFST\_DIAGN\_NUMBER]; //последние самые глубокие сообщения

GRBALPHABET\* lenta; //перекодированная(TS/NS) лента (из LEX)

short lenta\_position; //текущая позиция на ленте

short nrule; //номер текущего правила

short nrulechain; //номер текущей цепочки, текущего правила

short lenta\_size; //размер ленты

GRB::Greibach grebach; //грамматика Грейбах

Lexer::LEX lex; //результат работы лексического анализатора

MFSTSTSTACK st; //стек автомата

std::stack<MfstState> storestate; //стек для сохранения состояний

Mfst();

Mfst(

Lexer::LEX plex, //результат работы лексического анализатора

GRB::Greibach pgrebach //грамматика Грейбах

);

char\* getCSt(char\*buf); //получить содержимое стека

char\* getCLenta(char\* buf, short pos, short n = 25);//лента: n символов с pos

char\* getDiagnosis(short n, char\*buf);//получить n-ую строку диагностики или 0х00

bool savestate(const Log::LOG &log);//сохранить состояние автомата

bool reststate(const Log::LOG &log);//восстановить состояние автомата

bool push\_chain( //поместить уепочку правила в стек

GRB::Rule::Chain chain //цепочка правила

);

RC\_STEP step(const Log::LOG &log);//выполнить шаг автомата

bool start(const Log::LOG &log); //запустить автомат

bool savediagnois(

RC\_STEP pprc\_step); //код завершения шага

void printrules(const Log::LOG &log);//вывести последовательность правил

struct Deducation //вывод

{

short size; //количество шагов в выводе

short\* nrules; //номера правил нрамматики

short\* nrulechains;//номера цепочек правил грамматики (nrules)

Deducation() { size = 0; nrules = 0; nrulechains = 0; };

} deducation;

bool savededucation();}; //сохранить дерево вывода

Листинг 3 Структура грамматики Грейбах

struct Greibach //грамматика Грейбах

{

short size; //количество правил

GRBALPHABET startN; //стартовый символ

GRBALPHABET stbottomT;//дно стека

Rule \*rules; //множество правил

Greibach() { short size = 0; startN = 0; stbottomT = 0; rules = 0; };

Greibach(

GRBALPHABET pstartN, //стартовый символ

GRBALPHABET pstbootomT, //дно стека

short psize, //количество правил

Rule r, ... //правила

);

short getRule( //получить правило, возвращается номер правила или -1

GRBALPHABET pnn, //левый символ правила

Rule& prule //возвращаемое правило грамматики

);

Rule getRule(short n); }; //получить правило по номеру

Листинг 4 Разбор исходного кода синтаксическим анализатором

Шаг :Правило Входная лента Стек

0 :S->tfiPTS tfi(ti,ti)[nti;?i<i#w[i=i S$

1 : SAVESTATE: 1

1 : tfi(ti,ti)[nti;?i<i#w[i=i tfiPTS$

2 : fi(ti,ti)[nti;?i<i#w[i=i; fiPTS$

3 : i(ti,ti)[nti;?i<i#w[i=i;] iPTS$

4 : (ti,ti)[nti;?i<i#w[i=i;]r PTS$

5 :P->(E) (ti,ti)[nti;?i<i#w[i=i;]r PTS$

6 : SAVESTATE: 2

6 : (ti,ti)[nti;?i<i#w[i=i;]r (E)TS$

7 : ti,ti)[nti;?i<i#w[i=i;]r[ E)TS$

8 :E->ti,E ti,ti)[nti;?i<i#w[i=i;]r[ E)TS$

9 : SAVESTATE: 3

9 : ti,ti)[nti;?i<i#w[i=i;]r[ ti,E)TS$

10 : i,ti)[nti;?i<i#w[i=i;]r[i i,E)TS$

11 : ,ti)[nti;?i<i#w[i=i;]r[i= ,E)TS$

12 : ti)[nti;?i<i#w[i=i;]r[i=i E)TS$

13 :E->ti,E ti)[nti;?i<i#w[i=i;]r[i=i E)TS$

14 : SAVESTATE: 4

14 : ti)[nti;?i<i#w[i=i;]r[i=i ti,E)TS$

15 : i)[nti;?i<i#w[i=i;]r[i=i; i,E)TS$

16 : )[nti;?i<i#w[i=i;]r[i=i;] ,E)TS$

17 : TS\_NOK/NS\_NORULECHАIN

17 : RESSTATE

Листинг 4 (прод.) Разбор исходного кода синтаксическим анализатором

890 : SAVESTATE: 81

890 : i(i,i);] iF;]$

891 : (i,i);] F;]$

892 :F->(N) (i,i);] F;]$

893 : SAVESTATE: 82

893 : (i,i);] (N);]$

894 : i,i);] N);]$

895 :N->i i,i);] N);]$

896 : SAVESTATE: 83

896 : i,i);] i);]$

897 : ,i);] );]$

898 : TS\_NOK/NS\_NORULECHАIN

898 : RESSTATE

898 : i,i);] N);]$

899 :N->i,N i,i);] N);]$

900 : SAVESTATE: 83

900 : i,i);] i,N);]$

901 : ,i);] ,N);]$

902 : i);] N);]$

903 :N->i i);] N);]$

904 : SAVESTATE: 84

904 : i);] i);]$

905 : );] );]$

906 : ;] ;]$

907 : ] ]$

908 : $

909 : LENTA\_END

910 : ------>LENTA\_END

# Приложение Г

Листинг 1 Программная реализация механизма преобразования в ПОЛИЗ

bool setPolishNotation(IT::IdTable& idtable, Log::LOG& log, int lextable\_pos, ltvec& v)

{

vector < LT::Entry > result; //результирующий вектор

stack < LT::Entry > s; // стек для сохранения операторов

bool ignore = false; // флаг вызова функции

for (unsigned i = 0; i < v.size(); i++)

{

if (ignore) // вызов функции считаем подставляемым значением и заносим в результат

{

result.push\_back(v[i]);

if (v[i].lexema == LEX\_RIGHTTHESIS)

ignore = false;

continue;

}

int priority = getPriority(v[i]); // его приоритет

if (v[i].lexema == LEX\_LEFTHESIS || v[i].lexema == LEX\_RIGHTTHESIS || v[i].lexema == LEX\_PLUS || v[i].lexema == LEX\_MINUS || v[i].lexema == LEX\_STAR || v[i].lexema == LEX\_DIRSLASH)

{

if (s.empty() || v[i].lexema == LEX\_LEFTHESIS)

{s.push(v[i]);

continue;}

if (v[i].lexema == LEX\_RIGHTTHESIS)

{//выталкивание элементов до скобки

while (!s.empty() && s.top().lexema != LEX\_LEFTHESIS)

{result.push\_back(s.top());

s.pop();}

if (!s.empty() && s.top().lexema == LEX\_LEFTHESIS)

s.pop();

continue;

}

//выталкивание элем с большим/равным приоритетом while (!s.empty() && getPriority(s.top()) >= priority)

{

result.push\_back(s.top());

s.pop();

}

s.push(v[i]);

}

if (v[i].lexema == LEX\_LITERAL|| v[i].lexema == LEX\_ID) // идентификатор, идентификатор функции или литерал

{

if (idtable.table[v[i].idxTI].idtype == IT::IDTYPE::F || idtable.table[v[i].idxTI].idtype == IT::IDTYPE::S)

ignore = true;

result.push\_back(v[i]);// операнд заносим в вектор

}

if (v[i].lexema != LEX\_LEFTHESIS & v[i].lexema != LEX\_RIGHTTHESIS & v[i].lexema != LEX\_PLUS & v[i].lexema != LEX\_MINUS & v[i].lexema != LEX\_STAR & v[i].lexema != LEX\_DIRSLASH &v[i].lexema != LEX\_ID & v[i].lexema != LEX\_LITERAL)

{

Log::printError(log.stream, Error::GetError(1));

return false;

}

}

while (!s.empty()) { result.push\_back(s.top()); s.pop(); }

v = result;

return true;}

Листинг 2 Таблица идентификаторов после преобразования выражений в ПОЛИЗ

| N |СТРОКА В ТЛ| ТИП ИДЕНТИФИКАТОРА | ИМЯ | ЗНАЧЕНИЕ (ПАРАМЕТРЫ)

|0 | 2 | number function | min | P0:NUMBER | P1:NUMBER |

|1 | 5 | number parameter | minx |

|2 | 8 | number parameter | miny |

|3 | 13 | number variable | minres |0

|4 | 41 | proc function | stand | P0:STRING | P1:STRING |

|5 | 44 | string parameter | standa |

|6 | 47 | string parameter | standb |

|7 | 52 | number variable | standk |0

|8 | 56 | number LIB FUNC | length | P0:STRING |

|9 | 68 | string variable | standstr |[0]

|10 | 72 | string LIB FUNC | strcopy | P0:STRING

|11 | 91 | number variable | mainx |0

|12 | 93 | number literal | LTRL1 |9

|13 | 97 | number variable | mainy |0

|14 | 99 | number literal | LTRL2 |-9

|15 | 103 | string variable | mainstrx |[0]

|16 | 105 | string literal | LTRL3 |[4]Just

|17 | 109 | string variable | mainstry |[0]

|18 | 111 | string literal | LTRL4 |[6]string

|19 | 115 | number variable | maint |0

|20 | 133 | number variable | mainab |0

|21 | 135 | number literal | LTRL5 |3

|22 | 140 | number literal | LTRL6 |52

|23 | 148 | string literal | LTRL7 |[1]

|24 | 153 | number literal | LTRL8 |2

Листинг 3 Таблица лексем после преобразования выражений в ПОЛИЗ

| N | ЛЕКСЕМА | СТРОКА | ИНДЕКС В ТИ |

|0 | t | 1 | |

|1 | f | 1 | |

|2 | i | 1 | 0 |

|3 | ( | 1 | |

|4 | t | 1 | |

|5 | i | 1 | |

|6 | , | 1 | |

|7 | t | 1 | |

|8 | i | 1 | 2 |

|9 | ) | 1 | |

|10 | [ | 2 | |

|11 | n | 3 | |

|12 | t | 3 | |

|13 | i | 3 | 3 |

|14 | ; | 3 | |

|15 | ? | 4 | |

|16 | i | 4 | 1 |

|17 | < | 4 | |

|18 | i | 4 | 2 |

|19 | # | 4 | |

|20 | w | 5 | |

|21 | [ | 5 | |

|22 | i | 5 | 3 |

|23 | = | 5 | |

|24 | i | 5 | 1 |

|25 | ; | 5 | |

|26 | ] | 5 | |

|27 | r | 6 | |

|28 | [ | 6 | |

|29 | i | 6 | 3 |

|30 | = | 6 | |

|31 | i | 6 | 2 |

|32 | ; | 6 | |

|33 | ] | 6 | |

|34 | # | 6 | |

|35 | e | 7 | |

|36 | i | 7 | 3 |

|37 | ; | 7 | |

|38 | ] | 8 | |

|39 | p | 9 | |

|40 | f | 9 | |

|41 | i | 9 | 4 |

|42 | ( | 9 | |

|43 | t | 9 | |

|44 | i | 9 | 5 |

|45 | , | 9 | |

|46 | t | 9 | |

|47 | i | 9 |6 |

|48 | ) | 9 | |

|49 | [ | 10 | |

|50 | n | 11 | |

|51 | t | 11 | |

|52 | i | 11 |7 |

|53 | ; | 11 | |

|54 | i | 12 |7 |

|55 | = | 12 | |

|56 | i | 12 |8 |

|57 | ( | 12 | |

|58 | i | 12 |5 |

|59 | ) | 12 | |

|60 | ; | 12 | |

|61 | o | 13 | |

|62 | i | 13 |7 |

|63 | ; | 13 | |

|64 | ^ | 14 | |

|65 | ; | 14 | |

|66 | n | 15 | |

|67 | t | 15 | |

|68 | i | 15 |9 |

|69 | ; | 15 | |

|70 | i | 16 |9 |

|71 | = | 16 | |

|72 | i | 16 |10 |

|73 | ( | 16 | |

|74 | i | 16 |5 |

|75 | , | 16 | |

|76 | i | 16 |6 |

|77 | ) | 16 | |

|78 | ; | 16 | |

|79 | o | 17 | |

|80 | i | 17 |9 |

|81 | ; | 17 | |

|82 | ^ | 18 | |

|83 | ; | 18 | |

|84 | e | 19 | |

|85 | ; | 19 | |

|86 | ] | 20 | |

|87 | m | 21 | |

|88 | [ | 22 | |

|89 | n | 23 | |

|90 | t | 23 | |

|91 | i | 23 |11 |

|92 | = | 23 | |

|93 | l | 23 |12 |

|94 | ; | 23 | |

|95 | n | 24 | |

|96 | t | 24 | |

|97 | i | 24 |13 |

|98 | = | 24 | |

|99 | l | 24 |14 |

|100 | ; | 24 | |

|101 | n | 25 | |

|102 | t | 25 | |

|103 | i | 25 |15 |

|104 | = | 25 | |

|105 | l | 25 |16 |

|106 | ; | 25 | |

|107 | n | 26 | |

|108 | t | 26 | |

|109 | i | 26 |17 |

|110 | = | 26 | |

|111 | l | 26 |18 |

|112 | ; | 26 | |

|113 | n | 27 | |

|114 | t | 27 | |

|115 | i | 27 |19 |

|116 | ; | 27 | |

|117 | i | 28 |19 |

|118 | = | 28 | |

|119 | i | 28 |0 |

|120 | ( | 28 | |

|121 | i | 28 |11 |

|122 | , | 28 | |

|123 | i | 28 |13 |

|124 | ) | 28 | |

|125 | ; | 28 | |

|126 | o | 29 | |

|127 | i | 29 |19 |

|128 | ; | 29 | |

|129 | ^ | 30 | |

|130 | ; | 30 | |

|131 | n | 31 | |

|132 | t | 31 | |

|133 | i | 31 |20 |

|134 | = | 31 | |

|135 | l | 31 |21 |

|136 | ; | 31 | |

|137 | ? | 32 | |

|138 | i | 32 |20 |

|139 | ! | 32 | |

|140 | l | 32 |22 |

|141 | # | 32 | |

|142 | c | 33 | |

|143 | [ | 33 | |

|144 | o | 34 | |

|145 | i | 34 |20 |

|146 | ; | 34 | |

|147 | o | 35 | |

|148 | l | 35 |23 |

|149 | ; | 35 | |

|150 | i | 36 |20 |

|151 | = | 36 | |

|152 | i | 36 |20 |

|153 | l | 36 |24 |

|154 | + | 36 | |

|155 | l | 36 |24 |

|156 | \* | 36 | |

|157 | ; | 36 | |

|158 | ] | 37 | |

|159 | # | 37 | |

|160 | o | 38 | |

|161 | l | 38 |23

|162 | ; | 38 | |

|163 | o | 39 | |

|164 | i | 39 |20 |

|165 | ; | 39 | |

|166 | ^ | 40 | |

|167 | ; | 40 | |

|168 | i | 41 |4 |

|169 | ( | 41 | |

|170 | i | 41 |15 |

|171 | , | 41 | |

|172 | i | 41 |17 |

|173 | ) | 41 | |

|174 | ; | 41 | |

|175 | ] | 42 |

|

Листинг 4 Соответствие лексем исходному коду программы

1 | tfi[0](ti[1],ti[2])

2 | [

3 | nti[3];

4 | ?i[1]<i[2]#

5 | w[i[3]=i[1];]

6 | r[i[3]=i[2];]#

7 | ei[3];

8 | ]

9 | pfi[4](ti[5],ti[6])

10 | [

11 | nti[7];

12 | i[7]=i[8](i[5])l[9]+;

13 | ol[10];

14 | oi[7];

15 | ^;

16 | nti[11];

17 | i[11]=i[12](i[5],i[6]);

18 | ol[13];

19 | oi[11];

20 | ^;

22 | e;

23 | ]

24 | m

25 | [

26 | nti[14]=l[15];

27 | nti[16]=l[17];

28 | nti[18]=l[19];

29 | nti[20]=l[21];

30 | nti[22]=l[23];

31 | nti[24];

32 | ol[25];

33 | i[24]=i[26](i[22]);

34 | oi[24];

35 | ^;

36 | nti[27];

37 | i[27]=i[14]l[9]{;

38 | ol[28];

39 | oi[27];

40 | ^;

41 | nti[29];

42 | i[29]=i[0](i[14],i[16]);

43 | oi[29];

44 | ^;

45 | nti[30]=l[31];

46 | nti[32]=q;

47 | ?i[30]!l[33]#

48 | c[

49 | oi[30];

50 | ol[34];

51 | i[30]=i[30]l[35]+l[35]\*;

52 | ]#

53 | ol[34];

54 | oi[30];

55 | ^;

56 | i[4](i[18],i[20]);

57 | ]

# Приложение Д

Листинг 1 Результат генерации кода контрольного примера в Ассемблере

.586

.model flat, stdcall

includelib libucrt.lib

includelib kernel32.lib

includelib "../Debug/GenLib.lib"

ExitProcess PROTO:DWORD

.stack 4096

 outnum PROTO : DWORD

 outstr PROTO : DWORD

 strcopy PROTO : DWORD, : DWORD

 length PROTO : DWORD, : DWORD

 atoii  PROTO : DWORD

.const

newline byte 13, 10, 0

LTRL1 sdword 1

LTRL2 byte 'Len + 1: ', 0

LTRL3 byte 'Strcopy: ', 0

LTRL4 sdword 9

LTRL5 sdword -9

LTRL6 byte 'Just ', 0

LTRL7 byte 'string', 0

LTRL8 byte '70', 0

LTRL9 byte 'from string in number:', 0

LTRL10 byte 'Left shift:', 0

LTRL11 sdword 3

LTRL12 sdword 52

LTRL13 byte ' ', 0

LTRL14 sdword 2

.data

temp sdword ?

buffer byte 256 dup(0)

minres sdword 0

standk sdword 0

standstr dword ?

mainx sdword 0

mainy sdword 0

mainstrx dword ?

mainstry dword ?

mainstrz dword ?

maine sdword 0

mainresult sdword 0

maint sdword 0

mainab sdword 0

maind sdword 0

.code

;----------- min ------------

min PROC,

minx : sdword, miny : sdword

; --- save registers ---

push ebx

push edx

; ----------------------

mov edx, minx

cmp edx, miny

jl right1

jg wrong1

right1:

push minx

pop ebx

mov minres, ebx

jmp next1

wrong1:

push miny

pop ebx

mov minres, ebx

next1:

; --- restore registers ---

pop edx

pop ebx

; -------------------------

mov eax, minres

ret

min ENDP

;------------------------------

;----------- stand ------------

stand PROC,

standa : dword, standb : dword

; --- save registers ---

push ebx

push edx

; ----------------------

push standa

push offset buffer

call length

push eax

push LTRL1

pop ebx

pop eax

add eax, ebx

push eax

pop ebx

mov standk, ebx

push offset LTRL2

call outstr

push standk

call outnum

push offset newline

call outstr

push standb

push standa

push offset buffer

call strcopy

mov standstr, eax

push offset LTRL3

call outstr

push standstr

call outstr

push offset newline

call outstr

; --- restore registers ---

pop edx

pop ebx

; -------------------------

ret

stand ENDP

;------------------------------

;----------- MAIN ------------

main PROC

push LTRL4

pop ebx

mov mainx, ebx

push LTRL5

pop ebx

mov mainy, ebx

mov mainstrx, offset LTRL6

mov mainstry, offset LTRL7

mov mainstrz, offset LTRL8

push offset LTRL9

call outstr

push mainstrz

push offset buffer

call atoii

push eax

pop ebx

mov maine, ebx

push maine

call outnum

push offset newline

call outstr

push mainx

push LTRL1

pop ebx

pop eax

mov cl, bl

shl eax, cl

push eax

pop ebx

mov mainresult, ebx

push offset LTRL10

call outstr

push mainresult

call outnum

push offset newline

call outstr

push mainy

push mainx

call min

push eax

pop ebx

mov maint, ebx

push maint

call outnum

push offset newline

call outstr

push LTRL11

pop ebx

mov mainab, ebx

pop ebx

mov maind, ebx

mov edx, mainab

cmp edx, LTRL12

jnz cycle2

jmp cyclenext2

cycle2:

push mainab

call outnum

push offset LTRL13

call outstr

push mainab

push LTRL14

pop ebx

pop eax

add eax, ebx

push eax

push LTRL14

pop ebx

pop eax

imul eax, ebx

push eax

pop ebx

mov mainab, ebx

mov edx, mainab

cmp edx, LTRL12

jnz cycle2

cyclenext2:

push offset LTRL13

call outstr

push mainab

call outnum

push offset newline

call outstr

push mainstry

push mainstrx

call stand

push 0

call ExitProcess

main ENDP

end main