МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего профессионального образования

**"Южно-Уральский государственный университет"**

**(национальный исследовательский университет)**

**Факультет Вычислительной математики и информатики**

**Кафедра системного программирования**

**ОТЧЕТ   
по учебной практике**

бакалавра направления 09.03.02 "Информационные системы и

технологии"

|  |  |
| --- | --- |
|  | Выполнил:  студент группы ВМИ-105  Д.И. Булдаков  Проверил:  К.Ю. Никольская  старший преподаватель |

Челябинск-2016

Министерство образования и науки Российской Федерации

Южно-Уральский государственный университет

Кафедра системного программирования

УТВЕРЖДАЮ

Зав. кафедрой

системного программирования

\_\_\_\_\_\_\_\_\_\_\_\_\_ Л.Б. Соколинский

**ЗАДАНИЕ**

**по** **учебной практике**

1. **Цель работы**

Необходимо разработать распознаватель заданной символьной цепочки. Символьная цепочка задается с помощью формул Бэкуса-Наура.

1. **Исходные данные к работе**
2. *Йенсен К., Вирт Н.* Паскаль. Руководство пользователя и описание языка. М.: Компьютер, 1995.
3. *Льюис Ф., Розенкранц Д., Стирнз Р.* Теоретические основы проектирования компиляторов. М.: Мир, 1979.
4. **Перечень подлежащих разработке вопросов**
5. Выполнить анализ требований и разработать спецификации программы.
6. Провести проектирование программы.
7. Реализовать спроектированные модули.
8. Провести тестирование и отладку реализованных модулей.
9. **Сроки**

Дата выдачи задания: "27" июня 2016 г.

Срок сдачи законченной работы: "24" июля 2016 г.

**Руководитель:**

старший преподаватель \_\_\_\_\_\_\_\_\_\_\_\_\_ К.Ю. Никольская

подпись

**Задание принял к исполнению:**

\_\_\_\_\_\_\_\_\_\_\_\_\_ Д.И. Булдаков

подпись

**ОГЛАВЛЕНИЕ**

[1. Спецификация 4](#_Toc327919175)

[2. Проектирование 4](#_Toc327919176)

[2.1. Модульная структура 5](#_Toc327919177)

[2.2. Интерфейсы модулей 5](#_Toc327919178)

[3. Кодирование 7](#_Toc327919179)

[3.1. Структура текста программы 7](#_Toc327919180)

[3.2. Алгоритмы реализации модулей 7](#_Toc327919181)

[3.2.1. Блок транслитерации 7](#_Toc327919182)

[3.2.2. Лексический блок 8](#_Toc327919183)

[3.2.3. Синтаксический блок 12](#_Toc327919184)

[3.2.4. Блок идентификации ключевых слов 13](#_Toc327919185)

[3.3. Размер текста программы (в строках) 14](#_Toc327919186)

[4. Тестирование 14](#_Toc327919187)

[4.1. Автономное тестирование 14](#_Toc327919188)

[4.2. Комплексное тестирование 15](#_Toc327919189)

[Заключение 15](#_Toc327919190)

[Литература 15](#_Toc327919191)

# Спецификация

Необходимо разработать распознаватель заданной символьной цепочки. Символьная цепочка задается с помощью формул Бэкуса-Наура:

<цепочка>::=<оператор repeat-until>

<условие>::=<вызов подпрограммы>

<вызов подпрограммы>::=<идентификатор>**(** <список параметров> **)**;

<список параметров>::=<список идентификаторов>

<список идентификаторов>::=<идентификатор> |

<идентификатор> , <список идентификаторов>

<идентификатор>::=<буква>|

<идентификатор><буква>|

<идентификатор><цифра>

<буква>::=**A | B | C | D | E | F | ... | Z**

<цифра>::=**0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9**

<список параметров>::=<список идентификаторов>

<оператор>::=<вызов подпрограммы>

<список параметров>::=<элемент массива>

<элемент массива>::=<идентификатор>**[** <индексное выражение> **]**

<индексное выражение>::=<целая константа>

<целая константа>::=<целое со знаком> | <целое без знака>

<целое со знаком>::=<знак><целое без знака>

<знак>::= **+ | -**

<целое без знака>::=<цифра> | <целое без знака>

<список параметров>::=<целая константа>

<целая константа>::=<целое со знаком> | <целое без знака>

Помимо этого на цепочку накладывается следующее семантическое ограничение:

идентификатор, входящий в цепочку, не должен совпадать с ключевыми словами языка Pascal .

**Описание входных данных**

Цепочка записана в текстовом файле INPUT.TXT, который состоит из одной стро-ки. Длина цепочки не превышает 80 символов.

**Описание выходных данных**

Результат распознавания необходимо записать в текстовый файл OUTPUT.TXT в

одно из следующих сообщений: ACCEPT, если цепочка допустима, и REJECT, если цепочка недопустима.

**Примеры входных и выходных данных**

|  |  |
| --- | --- |
| **INPUT.TXT** | **OUTPUT.TXT** |
| repeat myproc(1) until myfunc(a); | ACCEPT |
| repeat myproc(+16) until myfunc(param1, param2); | ACCEPT |
| repeat myproc(-123) until myfunc(p1, p2, p3, p4); | ACCEPT |
| repeat begin(-123) until myfunc(p1, p2, p3, p4); | REJECT |
| repeat myproc(-123) until myfunc(p1, p2, p3, p4) | REJECT |
| repeat myproc(-123 par) until myfunc(p1, p2, p3, p4); | REJECT |
| repeat myproc(-123) until: myfunc(p1, p2, p3, p4); | REJECT |

# Проектирование

В данном разделе содержатся два подраздела: Модульная структура и Интерфейсы модулей. Вместе они описывают этап проектирования программы, до начала кодирования.

# Модульная структура

Программа

Синтаксический блок

**Блок транслитерации**

**Ввод/вывод данных**

**Лексический блок**

**Блок идентификации ключевых слов**

**Посимвольное разбиение строки и присваивание класса символа**

**Перевод символов в нижний регистр**

*Рисунок 1 – схема модульной структуры программы*

**Ввод/вывод данных -** ввода и вывода данных.

**Блок транслитерации** – Преобразование исходной символьной цепочки в двумерный вектор лексем вида ["символ цепочки"]["класс символа цепочки "]

**Перевод символов в нижний регистр** – Функция перевода всех символов строки в нижний регистр.

**Посимвольное разбиение строки** – Функция посимвольного разбиения строки, присваивающая символу его класс.

**Лексический блок** – Преобразование двумерного вектора лексем, полученного от блока транслитерации, в двумерный вектор лексем вида ["символ входного языка"]["класс символа входного языка"].

**Блок идентификации ключевых слов** – Блок, устанавливающий присутствуют ли в двумерном векторе лексем ключевые слова языка Паскаль.

**Перевод символов в верхний регистр** – Функция перевода названия ключевого слова в верхний регистр.

**Синтаксический блок** – Блок, устанавливающий соответствует ли вектор заданным формулам Бэкуса-Наура.

## 2.2. Интерфейсы модулей

Ввод данных:

Семантика: **string input(bool err);**

Входные данные: bool err – состояние ошибки. Если на вход подана пустая строка, будет выведено сообщение: "runtime error: empty string".

Выходные данные: string s – строка, считанная из файла input.txt.

Модуль возвращает строку из файла input.txt.

Вывод данных:

Семантика: **void output(bool result);**

Входные данные: bool result – переменная, в зависимости от которой в файл output.txt выводится ACCEPT или REJECT.

Выходные данные: –

Модуль в зависимости от значения булевой переменной result записывает в файл output.txt либо ACCEPT, либо REJECT.

Блок транслитерации:

Семантика: **string \*\* transliteration(string s, bool err);**

Входные данные: string s – строка, считанная из input.txt и bool err – состояние ошибки. Если будет встречен неописанный в блоке символ, будет выведено сообщение: "runtime error : unexpected symbol".

Выходные данные: возвращает двумерный массив типа string размера n x 2 вида ["символ цепочки"]["класс символа цепочки"], где n – длина исходной строки.

Например, символьную цепочку repeat myproc(1) until myfunc(a); блок транслитерации должен преобразовать в следующий двумерный вектор лексем:

|  |  |
| --- | --- |
| r | letter |
| e | letter |
| p | letter |
| e | letter |
| a | letter |
| t | space |
|  | letter |
| m | equal |
| y | number |
| p | number |
| r | dotcomma |
| o | letter |
| c | letter |
| ( | letter |
| 1 | letter |
| ) | letter |
|  | space |
| u | letter |
| n | equal |
| t | number |
| i | number |
| l | dotcomma |
|  | letter |
| m | letter |
| y | letter |
| f | letter |
| u | letter |
| n | space |
| c | letter |
| ( | equal |
| a | number |
| ) | number |
| ; | letter |

*Рисунок 1 - пример вектора блока транслитерации*

Перевод символов в нижний регистр:

Семантика: void tolower(string& str);

Входные данные: string& str – строка, считанная из input.txt

Выходные данные: –

Модуль проходит по строке и переводит все буквы в нижний регистр.

Лексический блок:

Семантика: **string \*\* lexicalBlock(string \*\* list, int & n);**

Входные данные:string \*\* list – двумерный массив, в котором находится строка, разбитая на символы с присвоенными им классами; int & n **–** длина входной строки.

Выходные данные: двумерный массив типа string размера n x 2

Блок отвечает за преобразование двумерного массива лексем, полученного от блока транслитерации, в двумерный массив лексем вида ["символ входного языка"]["класс символа входного языка"]. В рассматриваемом примере лексический блок должен выдать следующую цепочку лексем:

|  |  |
| --- | --- |
| repeat | id |
|  | space |
| myproc | id |
| ( | open |
| 1 | number |
| ) | close |
|  | space |
| until | id |
| myfunc | id |
| ( | open |
| a | number |
| ) | close |
| ; | dotcoma |

*Рисунок 2 - пример вектора лексического блока*

Блок идентификации ключевых слов:

Семантика: **bool keywordRecognize(string \*\* list, int & n);**

Входные данные: string \*\* list – двумерный массив, идентификаторы которого будут просматриваться на совпадения с ключевыми словами языка Паскаль; int & n – длина исходной строки.

Выходные данные: возвращает True если было обнаружено слово const в начале строки. Возвращает False если были обнаружены другие слова или слова const не было встречено в начале строки.

Блок устанавливает, присутствуют ли в двумерном массиве лексем ключевые слова языка Паскаль.

Синтаксический блок:

Семантика: **bool syntaxBlock(string \*\* list, int & n);**

Входные данные: **string \*\* list** – двумерный массив, содержащий лексемы и их классы; int & n – длина исходной строки.

Выходные данные: если конечный автомат перейдет в состояние НЕТ, то будет возвращено FALSE, иначе – TRUE.

Блок устанавливает, соответствует ли двумерный вектор заданным формулам Бэкуса-Наура. Данный блок использует для работы только классы символов входного языка. В рассматриваемом примере синтаксический блок, рассматривая цепочку вида KEYWORD\_CONST, SPACE, ID, EQUAL, NUMBER, SEMICOLON должен сообщить, что она синтаксически правильна.

# 3. Кодирование

## В данном пункте описан этап кодирования: структура текста программы, алгоритмы реализации модулей, размер текста программы. Алгоритмы реализации модулей поделены на три блока: блок транслитерации, лексический блок и синтаксический блок.

## 3.1. Структура текста программы

Practice.sln

lexicalBlock

transliteration

output

syntaxBlock

input

keywordRecognize

tolower

Описание всех функций в разделе ***2.2***.

# 3.2. Алгоритмы реализации модулей

В данном подразделе описаны детали реализации главных модулей программы.

### 3.2.1. Блок транслитерации

Все символы входной строки обрабатываются в два этапа: сначала переводятся в нижний регистр, затем разбиваются по классам, представленным в таблице 2.

**Таблица 2 - таблица транслитерации**

|  |  |
| --- | --- |
| **Символы** | **Класс лексемы** |
| a – z | letter |
| 0 - 9 | digit |
| ( | Open |
| ) | Close |
| , | Com |
| ; | Dotcoma |
|  | space |
| <другие символы> | err |

Если хотя бы одному символу был присвоен класс err, то в результате будет выведена ошибка "runtime error: unexpected symbol", а выполнение программы завершится.

### 3.2.2. Лексический блок

Построение обрабатывающего автомата начнем с построения “аналогичного” ему конечного распознавателя. В начале определим все возможные состояния распознавателя и их семантику в таблице 3.

**Таблица 3 – состояния конечного распознавателя лексического блока**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Состояние** | **Семантика** |
| 1 | НАЧ | Момент до начала обработки цепочки, либо чтение пробелов в начале строки. |
| 2 | КЛСЛОВО | Чтение ключевого слова. |
| 3 | ПРОБЕЛ1 | Чтение пробелов между ключевым словом и идентификатором. |
| 4 | ИД | Чтение идентификатора. |
| 5 | ПРОБЕЛ2 | Чтение пробелов между идентификатором и знаком равно. |
| 6 | РАВНО | Считан знак **равно**. |
| 7 | ПРОБЕЛ3 | Чтение пробелов между знаком равно и числом или знаком. |
| 8 | ЗНАК | Считан знак **+** или **-**. |
| 9 | ЧИСЛО | Чтение числа. |
| 10 | ДОЛЛАР | Считан знак **$**. |
| 11 | ЧИСЛОБ | Чтение 16-ричного числа. |
| 12 | ПРОБЕЛ4 | Чтение пробелов после числа. |
| 13 | ТЧКЗПТ | Считан знак **точка с запятой**. |
| 14 | ПРОБЕЛ5 | Чтение пробелов после **точки с запятой**. |
| 15 | E | Ошибка. |

Начальное состояние НАЧ. Допустимые состояния – ТЧКЗПТ и ПРОБЕЛ5. Конечный распознаватель, построенный на основе полученных состояние, приведен в таблице 4.

**Таблица 4 – конечный распознаватель лексического блока**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | letter | 16\_letter | digit | space | equal | semicolon | dollar | sign |  |
| **НАЧ** | КЛСЛОВО | КЛСЛОВО | Е | НАЧ | Е | Е | Е | Е | 0 |
| **КЛСЛОВО** | КЛСЛОВО | КЛСЛОВО | Е | ПРОБЕЛ1 | Е | Е | Е | Е | 0 |
| **ПРОБЕЛ1** | ИД | ИД | Е | ПРОБЕЛ1 | Е | Е | Е | Е | 0 |
| **ИД** | ИД | ИД | ИД | ПРОБЕЛ2 | РАВНО | Е | Е | Е | 0 |
| **ПРОБЕЛ2** | Е | Е | Е | ПРОБЕЛ2 | РАВНО | Е | Е | Е | 0 |
| **РАВНО** | Е | Е | ЧИСЛО | ПРОБЕЛ3 | Е | Е | ДОЛЛАР | ЗНАК | 0 |
| **ПРОБЕЛ3** | Е | Е | ЧИСЛО | ПРОБЕЛ3 | Е | Е | ДОЛЛАР | ЗНАК | 0 |
| **ЗНАК** | Е | Е | ЧИСЛО | Е | Е | Е | Е | Е | 0 |
| **ЧИСЛО** | Е | Е | ЧИСЛО | Е | Е | ТЧКЗПТ | Е | Е | 0 |
| **ДОЛЛАР** | Е | ЧИСЛОБ | ЧИСЛОБ | Е | Е | Е | Е | Е | 0 |
| **ЧИСЛОБ** | Е | ЧИСЛОБ | ЧИСЛОБ | Е | Е | ТЧКЗПТ | Е | Е | 0 |
| **ПРОБЕЛ4** | Е | Е | Е | ПРОБЕЛ4 | Е | ТЧКЗПТ | Е | Е | 0 |
| **ТЧКЗПТ** | Е | Е | Е | ПРОБЕЛ5 | Е | Е | Е | Е | 0 |
| **ПРОБЕЛ5** | Е | Е | Е | ПРОБЕЛ5 | Е | Е | Е | Е | 0 |
| **E** | Е | Е | Е | Е | Е | Е | Е | Е | 1 |

В полученном конечном распознавателе нет недостижимых состояний. Идентичными (а значит эквивалентными) состояниями являются РАВНО и ПРОБЕЛ3, ТЧКЗПТ и ПРОБЕЛ5. В дальнейшем будем обозначать их как РАВНО и ТЧКЗПТ соответственно. Редукция данного конечного распознавателя приведена в таблице 5.

**Таблица 5 – редукция конечного распознавателя лексического блока**

|  |  |  |
| --- | --- | --- |
| **Шаг** | **Результат (блоки состояний)** | **Действия** |
| 0 | Р0={НАЧ, КЛСЛОВО, ПРОБЕЛ1, ИД, ПРОБЕЛ2, РАВНО, ЗНАК, ЧИСЛО, ДОЛЛАР, ЧИСЛОБ, ПРОБЕЛ4, ТЧКЗПТ, Е} | Разбиваем Р0 на два блока: допустимые и отвергающие. |
| 1 | Р={НАЧ, КЛСЛОВО, ПРОБЕЛ1, ИД, ПРОБЕЛ2, РАВНО, ЗНАК, ЧИСЛО, ДОЛЛАР, ЧИСЛОБ, ПРОБЕЛ4, Е}  Р={ТЧКЗПТ} | Разбиваем Р по входу equal. |
| 2 | Р={НАЧ, КЛСЛОВО, ПРОБЕЛ1, РАВНО, ЗНАК, ЧИСЛО, ДОЛЛАР, ЧИСЛОБ, ПРОБЕЛ4, Е}  Р={ИД, ПРОБЕЛ2}  Р={ТЧКЗПТ} | Разбиваем Р по входу semicolon. |
| 3 | Р={НАЧ, КЛСЛОВО, ПРОБЕЛ1, РАВНО, ЗНАК, ДОЛЛАР, Е}  Р={ЧИСЛО, ЧИСЛОБ, ПРОБЕЛ4}  Р={ИД, ПРОБЕЛ2}  Р={ТЧКЗПТ} | Разбиваем по входу Р 16\_letter.  Разбиваем по входу Р 16\_letter.  Разбиваем по входу Р 16\_letter |
| 4 | Р={НАЧ, КЛСЛОВО}  Р={ПРОБЕЛ1}  Р={ИД}  Р={РАВНО, ЗНАК, Е}  Р={ЧИСЛО, ПРОБЕЛ4}  Р={ДОЛЛАР}  Р={ЧИСЛОБ}  Р={ПРОБЕЛ2}  Р={ТЧКЗПТ} | Разбиваем по входу Р по space.  Разбиваем по входу Р по space.  Разбиваем по входу Р по space. |
| 5 | Р={НАЧ}  Р={КЛСЛОВО}  Р={ПРОБЕЛ1}  Р={ИД}  Р={ЗНАК, Е}  Р={РАВНО}  Р={ЧИСЛО}  Р={ПРОБЕЛ4}  Р={ДОЛЛАР}  Р={ЧИСЛОБ}  Р={ПРОБЕЛ2}  Р={ТЧКЗПТ} | Разбиваем по входу Р по digit. |
| 6 | Р={НАЧ }  Р={КЛСЛОВО}  Р={ПРОБЕЛ1}  Р={ИД}  Р={ДОЛЛАР}  Р={РАВНО}  Р={ЗНАК}  Р={Е}  Р={ЧИСЛО}  Р={ПРОБЕЛ4}  Р={ЧИСЛОБ}  Р={ПРОБЕЛ2}  Р={ТЧКЗПТ} |  |

Проведенные преобразования доказывают, что конечный распознаватель, полученный после замены эквивалентных состояний, является минимальным. Данный распознаватель приведен в таблице 6.

**Таблица 6 – минимальный конечный распознаватель лексического блока.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | letter | 16\_letter | digit | space | equal | semicolon | dollar | sign |  |
| **НАЧ** | КЛСЛОВО | КЛСЛОВО | Е | НАЧ | Е | Е | Е | Е | 0 |
| **КЛСЛОВО** | КЛСЛОВО | КЛСЛОВО | Е | ПРОБЕЛ1 | Е | Е | Е | Е | 0 |
| **ПРОБЕЛ1** | ИД | ИД | Е | ПРОБЕЛ1 | Е | Е | Е | Е | 0 |
| **ИД** | ИД | ИД | ИД | ПРОБЕЛ2 | РАВНО | Е | Е | Е | 0 |
| **ПРОБЕЛ2** | Е | Е | Е | ПРОБЕЛ2 | РАВНО | Е | Е | Е | 0 |
| **РАВНО** | Е | Е | ЧИСЛО | РАВНО | Е | Е | ДОЛЛАР | ЗНАК | 0 |
| **ЗНАК** | Е | Е | ЧИСЛО | Е | Е | Е | Е | Е | 0 |
| **ЧИСЛО** | Е | Е | ЧИСЛО | ПРОБЕЛ4 | Е | ТЧКЗПТ | Е | Е | 0 |
| **ДОЛЛАР** | Е | ЧИСЛОБ | ЧИСЛОБ | Е | Е | Е | Е | Е | 0 |
| **ЧИСЛОБ** | Е | ЧИСЛОБ | ЧИСЛОБ | ПРОБЕЛ4 | Е | ТЧКЗПТ | Е | Е | 0 |
| **ПРОБЕЛ4** | Е | Е | Е | ПРОБЕЛ4 | Е | ТЧКЗПТ | Е | Е | 0 |
| **ТЧКЗПТ** | Е | Е | Е | ТЧКЗПТ | Е | Е | Е | Е | 0 |
| **E** | Е | Е | Е | Е | Е | Е | Е | Е | 1 |

Теперь по полученному конечному распознавателю составим обрабатывающий автомат. Для этого введем примитивные процедуры, описанные в таблице 7.

**Таблица 7 – примитивные процедуры обрабатывающего автомата лексического блока**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Процедура** | **Семантика** |
| 1 | *Да* | Остановить обработку и допустить цепочку. |
| 2 | *Нет* | Остановить обработку и отвергнуть цепочку. |
| 3 | *Обработать* | Добавить входной символ к значению текущей лексемы. |
| 4 | *Проверить id* | Проверить, совпадает ли текущая лексема с ключевыми словами языка Pascal. Если нет – продолжить обработку, если да – вызвать процедуру *Нет.* |
| 5 | *Проверить key\_word* | Проверить, является ли текущая лексема ключевым словом языка Pascal. Если да – продолжить обработку, если нет – вызвать процедуру *Нет.* |
| 6 | *Сохранить* | Сохранить текущую лексему в памяти. |
| 7 | *Лексема (класс)* | Перейти к рассмотрению следующей лексемы с заданным классом |

*Проверить id* и *Проверить key\_word* проверяются с помощью вызова блока идентификации ключевых слов.

Для удобства определим процедуры переходов, которые будем обозначать цифрами. Данные процедуры приведены 8. В случае, если встречается несколько процедур, то они выполняются последовательно и лишь потом происходит переход.

**Таблица 8 – процедуры переходов обрабатывающего автомата лексического блока**

|  |  |
| --- | --- |
| **Действие** | **Семантика** |
| 1 | Обработать; |
| 2 | Сохранить; |
| 3 | Лексема (key\_word);  Обработать; |
| 4 | Лексема (id);  Обработать; |
| 5 | Лексема (equal);  Обработать; |
| 6 | Лексема (number);  Обработать; |
| 7 | Лексема (semicolon);  Обработать; |
| 8 | Проверить id; |
| 9 | Проверить key\_word; |

Теперь можно на основе минимального конечного распознавателя составить соответствующий обрабатывающий автомат, он приведен в таблице 9.

**Таблица 9 – обрабатывающего автомата лексического блока.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | letter | 16\_letter | digit | space | equal | semicolon | dollar | sign |
| **НАЧ** | **3**  КЛСЛОВО | **3**  КЛСЛОВО |  | НАЧ |  |  |  | Е |
| **КЛСЛОВО** | **1**  КЛСЛОВО | **1**  КЛСЛОВО |  | **2, 9**  ПРОБЕЛ1 |  |  |  | Е |
| **ПРОБЕЛ1** | **4**  ИД | **4**  ИД |  | ПРОБЕЛ1 |  |  |  | Е |
| **ИД** | **1**  ИД | **1**  ИД | **1**  ИД | **8, 2, 5**  ПРОБЕЛ2 | **8, 2**  РАВНО |  |  | Е |
| **ПРОБЕЛ2** |  |  |  | **5**  ПРОБЕЛ2 | РАВНО |  |  | Е |
| **РАВНО** |  |  | **6**  ЧИСЛО | РАВНО |  |  | **6**  ДОЛЛАР | **6**  ЗНАК |
| **ЗНАК** |  |  | **1**  ЧИСЛО |  |  |  |  | Е |
| **ЧИСЛО** |  |  | **1**  ЧИСЛО | ПРОБЕЛ4 |  | **7**  ТЧКЗПТ |  | Е |
| **ДОЛЛАР** |  | **1**  ЧИСЛОБ | **1**  ЧИСЛОБ |  |  |  |  | Е |
| **ЧИСЛОБ** |  | **1**  ЧИСЛОБ | **1**  ЧИСЛОБ | ПРОБЕЛ4 |  | **7**  ТЧКЗПТ |  | Е |
| **ПРОБЕЛ4** |  |  |  | ПРОБЕЛ4 |  | **7**  ТЧКЗПТ |  | Е |
| **ТЧКЗПТ** |  |  |  | ТЧКЗПТ |  |  |  | Е |

Все пустые клетки обозначают вызов примитивной процедуры *НЕТ.*

### 3.2.3. Синтаксический блок

Составим конечный распознаватель, опираясь на исходные формулы Бэкуса-Наура и полученные в лексическом блоке лексемы: id, key\_word, equal, number, semicolon. Данный распознаватель приведен в таблице 10, он имеет только одно допускающее состояние ТЧКЗПТ.

**Таблица 10 – конечный распознаватель синтаксического блока**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | key\_word | id | equal | number | semicolon | E |  |
| **НАЧ** | КЛСЛОВО | Е | Е | Е | Е | Е | 0 |
| **КЛСЛОВО** | Е | ИД | Е | Е | Е | Е | 0 |
| **ИД** | Е | Е | РАВНО | Е | Е | Е | 0 |
| **РАВНО** | Е | Е | Е | ЧИСЛО1 | Е | Е | 0 |
| **ЧИСЛО** | Е | Е | Е | Е | ТЧКЗПТ | Е | 0 |
| **ТЧКЗПТ** | Е | Е | Е | Е | Е | Е | 1 |
| **Е** | Е | Е | Е | Е | Е | Е | 0 |

Данный распознаватель является минимальным, это можно доказать методом разбиений, последовательно разбив Р0 на допустимые и отвергающие состояние, а допустимый блок по входам semicolon, number, equal, id, key\_word.

Для составления обрабатывающего автомата вводим примитивные процедуры, приведенные в таблице 11.

**Таблица 11 – примитивные процедуры обрабатывающего автомата синтаксического блока**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Процедура** | **Семантика** |
| 1 | Да | Остановить обработку и допустить цепочку. |
| 2 | Нет | Остановить обработку и отвергнуть цепочку. |
| 3 | Проверить key\_word | Проверить, совпадает ли текущая лексема с ключевым словом const. Если да, то продолжить обработку. Иначе остановить обработку и отвергнуть цепочку. |

Обрабатывающий автомат, построенный на основе минимального конечного распознавателя, приведен в таблице 13.

**Таблица 12 – обрабатывающий автомат синтаксического блока**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | key\_word | id | equal | number | semicolon | ˧ |
| **НАЧ** | *Проверить key\_word*  КЛСЛОВО |  |  |  |  |  |
| **КЛСЛОВО** |  | ИД |  |  |  |  |
| **ИД** |  |  | РАВНО |  |  |  |
| **РАВНО** |  |  |  | ЧИСЛО1 |  |  |
| **ЧИСЛО** |  |  |  |  | ТЧКЗПТ |  |
| **ТЧКЗПТ** |  |  |  |  |  | ДА |

Все пустые клетки обозначают вызов примитивной процедуры *НЕТ*.

Реализация данного блока почти полностью идентична реализация лексического блока.

### 3.2.4. Блок идентификации ключевых слов

Реализация функции bool keywordRecognize(string \*\* list, int & n);

Подключается файл keywords.txt, содержащий ключевые слова языка Паскаль.

С помощью цикла for в массиве, содержащем ["лексема"]["класс лексемы"] находим класс лексемы **идент.**

Далее соответствующую классу лексему сравниваем со всеми идентификаторами в файлеkeyword.txt. Если совпадение есть, то функция заменяет возвращает False, иначе продолжится поиск идентификаторов. Кроме этого проверяется, что лексема класса **кл\_слово** – слово const. Как только все идентификаторы будут просмотрены, функция оканчивает работу.

## 3.3. Размер текста программы (в строках)

Всего программа занимает 600 строк.

# 4. Тестирование

Данный раздел содержит протоколы автономного и комплексного тестирования различных блоков распознавателя цепочки.

## 4.1. Автономное тестирование

Часть протоколов тестирования для блоков транслитерации, лексического блока, блока идентификации ключевых слов и синтаксического блока приведены соответственно в таблицах 13, 14, 15 и 16.

**Таблица 13 – часть протокола автономного тестирования блока транслитерации**

|  |  |  |
| --- | --- | --- |
| **Входные данные** | **Данные в векторе** | **Верно ли?** |
| const bc= 8; | (c, 16\_letter), (o, letter), (n, letter), (s, letter), (t, letter), ( , space), (b, 16\_letter), (c, 16\_letter), (=, equal), ( , space), (8, digit), (;, semicolon) | **Да** |
| 2045 | (2, digit), (0, digit), (4, digit), (5, digit) | **Да** |
| 32 – 4 = 28 | (3, digit), (2, digit), ( , space), (-, sign), (4, digit), ( , space), (4, digit), ( , space), (=, equal), ( , space), (2, digit), (8, digit) | **Да** |
| й=34; | runtime error: unexpected symbol | **Да** |
| if a | (i,, letter), (f, 16\_letter), ( , space), (a, 16\_letter) | **Да** |

**Таблица 14 – часть протокола автономного тестирования лексического блока**

|  |  |  |
| --- | --- | --- |
| **Входные данные** | **Выходные данные** | **Верно ли?** |
| (c, 16\_letter), (o, letter), (n, letter), (s, letter), (t, letter), ( , space), (b, 16\_letter), (c, 16\_letter), (=, equal), ( , space), (8, digit), (;, semicolon) | (const, кл\_слово), (bc, ид), (=, равно), (8, число), (;, тчкзпт) | **Да** |
| (i,, letter), (f, 16\_letter), ( , space), (a, 16\_letter) | (if, кл\_слово), (a, id) | **Да** |
| (c, 16\_letter), (o, letter), (n, letter), (s, letter), (t, letter), ( , space), (a, 16\_letter) | (const, кл\_слово), (a, ид) | **Да** |

**Таблица 15 – часть протокола автономного тестирования блока идентификации ключевых слов**

|  |  |  |
| --- | --- | --- |
| **Входные данные** | **Выходные данные** | **Верно ли?** |
| const | false | Да |
| fgR | true | Да |
| repeat | false | Да |
| ringf | true | Да |

## 4.2. Комплексное тестирование

В таблице 16 приведена часть протокола тестирования головного модуля программы.

**Таблица 16 – часть протокола тестирования головного модуля**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№ п/п** | **Входные данные** | **Выходные данные** | **Действительный результат** | **Тест пройден?** |
| 1 | const g = $ab3d; | **ACCEPT** | **ACCEPT** | **Да** |
| 2 | const re=-12; | **ACCEPT** | **ACCEPT** | **Да** |
| 3 | consT xabe=456; | **ACCEPT** | **ACCEPT** | **Да** |
| 4 | Const QWErTy = $a1b2 ; | **ACCEPT** | **ACCEPT** | **Да** |
| 5 | const FI= $998; | **ACCEPT** | **ACCEPT** | **Да** |
| 6 | conft N=$abc; | **REJECT** | **REJECT** | **Да** |
| 7 | const pretty = $+7; | **REJECT** | **REJECT** | **Да** |
| 8 | const end=13; | **REJECT** | **REJECT** | **Да** |
| 9 | ConsTlise=24556; | **REJECT** | **REJECT** | **Да** |

# Заключение

Данная работа посвящена разработке распознавателя символьной цепочки, заданной формулами Бэкуса-Наура. В ходе выполнения работы были пройдены все основные этапы разработки программного обеспечения: анализ, написание спецификации, проектирование, разработка алгоритмов, кодирование, тестирование и сопровождение. Каждому этапу, за исключением сопровождения, в данной работе уделен отдельный пункт. В данных пунктах описана проделанная в каждом случае работа и ее результаты.

# Литература

1. Льюис Ф., Розенкранц Д., Стирнз Р. Теоретические основы проектирования компиляторов. - М.: Мир, 1979.

2. Йенсен К., Вирт Н. Паскаль. Руководство пользователя и описание языка. - М.: Компьютер, 1995.