Министерство науки и высшего образования Российской Федерации

Новосибирский государственный технический университет

Кафедра вычислительной техники

Отчет по лабораторной работе № 2

по дисциплине «Теория формальных языков и компиляторов»

на тему «Лексика языков программирования. Конечные автоматы без памяти для обнаружения слов в тексте программы»

Студент: Шалаев И. А.

Вариант: 41111111

Группа: АВТ-709

Преподаватель: Малявко А.А.

Новосибирск, 2020

# Цель работы

Изучение конечных автоматов (КА) без памяти, способов определения КА – канонического, графового и табличного, методов построения недетерминированного КА по системе регулярных выражений, методов эквивалентных преобразований недетерминированных КА в оптимальные полностью определенные КА – лексические акцепторы.

# Постановка задачи

1. Используя пакет ВебТрансЛаб, освоить:

- создание лексических правил на языке регулярных выражений (РВ);

- использование операций «+, \*, ?, конкатенации и выбора» языка РВ для построения сложных регулярных выражений;

- преобразование простой системы РВ в одноавтоматный лексический акцептор;

- добавление правил и действий в систему РВ для построения мультиавтоматного лексического акцептора;

2. Разработать (доработать разработанный при выполнении работы №1) фрагмент системы регулярных выражений для всех (или выбранной самостоятельно части) групп слов языка, определенного заданием на курсовую работу. Построить по этому фрагменту:

- программный модуль, управляемый графом состояний и переходов;

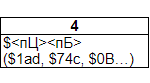
- программный модуль, управляемый таблично;

3. Изучить структуру программных модулей, построенных ВебТрансЛабом, изучить алгоритмы работы лексического акцептора для графового и табличного способов реализации КА, сравнить реализации конечных автоматов, управляемых различными способами, между собой, оценить их затраты времени в шагах на распознавание слов;

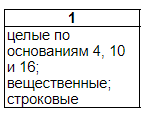
4. Изучить по тексту программного модуля способ реализации вызова действий, определенных в лексических правилах и алгоритм работы формирователя лексем.

5. Проверить функционирование конечных автоматов, построенных ВебТрансЛабом (подготовить тестовый пример программы на языке, заданном на курсовую работу, запустить каждый автомат на выполнение, протрассировать вручную работу лексического акцептора в графовой и табличной реализации, убедиться в работоспособности автоматов, в противном случае – доработать систему РВ и добиться правильного функционирования лексического акцептора).

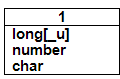
Вариант:



*Рис. 1 – Идентификаторы*



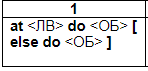
*Рис. 2 – Константы*



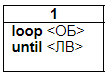
*Рисунок 3 – Типы*



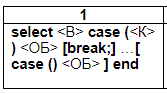
*Рисунок 4 – Оператор присваивания*



*Рисунок 5 – Условный оператор*



*Рисунок 6 – Оператор цикла*



*Рисунок 7 – Оператор переключателя*

# Ход работы

**Тестоваяпрограмма*:***

*//57. Найти минимальное число, большее заданного, которое может быть представлено*

*//как сумма степеней 2, 3, 4 и 5 различных простых чисел (наименьшее такое число*

*//есть 287 = 72+53+34+25, следующее – 359 = 112+53+34+25, …)*

*#include <stdio.h>*

*#include <math.h>*

*#include <ctime>*

*const long $1SIZE = 21;*

*long $0f(long $0n) {*

*const long $1SIZE = 21;*

*long $0simple[$1SIZE] = { 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31,*

*37, 41, 43, 47, 53, 59, 61, 67, 71,*

*73};*

*long $0i = 0;*

*long $0j = 0;*

*long $0k = 0;*

*long $0m = 0;*

*long $0min = $0pow($0simple[$1SIZE], 4);*

*loop {*

*$0i++*

*Loop {*

*{$0j++*

*loop {*

*$0k++*

*loop {*

*$0m++*

*long\_u $0temp = $0pow($0simple[$0i], 2) + $0pow($0simple[$0j], 3) + $0pow($0simple[$0k], 4) + $0pow($0simple[$0m], 5);*

*//$0printf("%d^2 + %d^3 + %d^4 + %d^5 = %d\n", $0simple[$0i], $0simple[$0j], $0simple[$0k], $0simple[$0m], $0temp);*

*at (($0temp > $0n) && ($0temp < $0min)) do*

*$0min = $0temp;*

*} until ($0m < $1SIZE)*

*} until ($0k < $1SIZE)*

*} until ($0j < $1SIZE)*

*} until ($0i < $1SIZE)*

*return $0min;*

*}*

*void $0do(){*

*$0printf("%s", "Enter N: ");*

*long $0n;*

*$0scanf("%d", &$0n);*

*number $0starttime = $0clock();*

*$0printf("N: %d\nResult: %d \n", $0n ,$0f($0n));*

*number $0endtime = $0clock();*

*number $0searchtime = $0endtime - $0starttime;*

*$0printf("%s%d ms\n", "Time: ", $0searchtime);*

*long $0entry() {*

*$0i = 0;*

*long $0n;*

*$0scanf("%d", &$0n);*

*loop*

*select ($0n)*

*case(1) {*

*$0do();*

*} break;*

*case(2) {*

*$0i =1;*

*} break;*

*end until ($0i == 0)*

*return 0;*

}

**Основные понятия**

**Произвольное выражение** – это совокупность идентификаторов, знаков математических операций и круглых скобок подчиняемая математическим законам.

**Логическое выражение** – это выражение результат, которого можно интерпретировать как истину или ложь.

**Константа** – это число, символ или строка символов, которая в программе используется, как значение. Значение константы нельзя изменить.

**Оператор** – один выполняемый оператор (присваивания, условный оператор, оператор цикла, переключатель)

**Блок** – это набор идущих подряд операторов, обрамлённый фигурными скобками.

**Функция** – это фрагмент программного кода (подпрограмма), к которому можно обратиться из другого места программы.

**Возвращаемое значение** – это значение, которое вернет функция после выполнения. Тип возвращаемого выражения должен совпадать с типом функции.

**Файл**

В файле с исходным кодом программы располагаются команды препроцессора и функции. Причем только в таком порядке. Функции должны располагаться последовательно, то есть функция должна располагаться над вызвавшей ее функцией. Последней функцией должна быть функция *$0entry*.

Переменные могут быть описаны либо в командах препроцессора, либо внутри функций.

Функции *$0entry* может и не быть файле, но тогда этот файл для выполнения функций в нем должен быть подключен к файлу, где эта функция есть.

**Расположение**

Программа может располагаться в нескольких файлах.

Подключение дополнительных файлов происходит путем указания их имени специальной командой *#include <имя\_файла>.*

Включаемый файл должен находится в директории включаемых файлов, указанной в настройках проекта.

Подключаемые файлы также могут иметь директивы *#include.*

**Команды препроцессора**

Существуют команды препроцессора, которые будут выполнены до начала работы функции *$0entry*. Они начинаются с символа ‘#’, такие как *#define* или *#include.*

Пример:

*#include <stdio.h>*

*#define $0pi 3.14;*

**Точка входа**

Точка входа — адрес в оперативной памяти, с которого начинается выполнение программы, адрес, по которому хранится первая команда программы.

Точка входа определяется ключевым словом *$0entry.*

То есть программа всегда начинает свою работу с выполнения функции *$0entry* и заканчивает свое выполнение с окончанием операторов в ней.

Начать выполнение программы без этой функции невозможно.

В программе допускается только одна функция *$0entry*, иначе возникнет ошибка.

**Идентификаторы**

Идентификаторы всегда начинаются с символа ‘$’, далее идет непустая последовательность десятичных цифр, далее непустая последовательность английских букв.

Прописные английские буквы: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Строчные английские буквы: a b c d e f g h i j k l m n o p q r s t u v w x y z

Десятичные цифры: 0 1 2 3 4 5 6 7 8

Пример: *$1337value;*

Компилятор обрабатывает буквы в разном регистре, как разные символы. То есть идентификаторы *$0n* и *$0N* будут восприняты компилятором как разные.

**Переменные**

Переменная може быть объявлена в любом месте функции и будет существовать только в ней.

Если переменная объявлена в функции *$0Fun*, то в функции *$0AnFun* ее не будет видно, и в функции *$0AnFun* может быть объявлена переменная с таким же именем.

В команде препроцессора переменную можно объявить, как константную и доступную в любом месте программы

Переменные должны быть объявлены и инициализированы до их использования.

Объявление: *long $0n;*

Инициализация: *$0n = 155;*

Возможно инициализировать переменную при объявлении:

*long $0n = 155;*

**Типы переменных**

*long* – целочисленный тип данных.

*long\_u* – без знаковое целое число.

*number* – число с плавающей запятой.  
*char* – символьный тип данных.

Существует тип *void* (аналогично языку С), но используется только для создания функций.

**Константы**

Константы тоже имеют область видимости. В случае определения константы в теле функции/цикла область ее видимости - функция/цикл, в котором она была определена.

**Целые константы**

Целые константы могут быть десятичными, двоичными или восьмеричными:  
*6* – Константа по основанию 10  
*4х02* – Константа по основанию 4

*16х1F* – Константа по основанию 16

**Вещественные константы**

Вещественные константы могут быть только десятичными:  
*12.2  
700.0  
1.9999*

**Символьные константы**

Символьные константы состоят из одного любого символа, заключенного в одинарные кавычки:

*‘a’  
‘0’  
‘<’***Строковые константы**

Строковые константы состоят из последовательности любых символов, возможно пустой, заключенной в двойные кавычки:

*“It is works”*

Можно использовать управляющие последовательности внутри таких констант:

*\n* – для перехода на новую строку

*\t* – для табуляции  
*“Hello\nworld”*

*“Hello\tworld”*

**Функции**

Функция определяется следующим образом:  
*<тип> <идентификатор> ([<аргументы>]) {[<Оператор\_или\_блок>]}*

Тело функции – набор операторов между открывающей и закрывающей фигурной скобкой.

Если функция принимает аргументы, то тогда они должны быть объявлены в поле *<аргументы>.* Переданные в функцию аргументы будут существовать только ней так же, как и созданные в ней переменные.

Если функция имеет возвращаемое значение, то тогда в ней должно быть указано ключевое слово *return* которое передаст управление оператору, следующему за вызовом этой функции. Оно должно быть указано во всех местах возможного выхода из функции. В функциях с типом *void*, не требуется указания ключевого слово *return*.

Пример:

*long $0faktorial(long $0numb)*

*{*

*long $0rezult = 1;*

*int $0i = 1*

*loop*

*$0rezult = $0rezult \* $0i;*

*$0i++;*

*until ($0i <= $0numb)*

*return $0rezult;*

*}*

Тип аргументов, переданных в функцию, должен совпадать с типом аргументов, указанных при объявлении функции.

Пример вызова вышеописанной функции:

*Long $0n = 5;*

*$0faktorial($0n);*

**Выполняемые операторы**

Присваивание:

*<Идентификатор>=<Произвольное\_выражение>*

Пример:

*$0n = 5*

Условный оператор:

*at (<Логическое\_выражение>) do <Оператор\_или\_блок> else do <Оператор\_или\_блок>]*

Пример:

*at (($0temp > $0n)  && ($0temp < $0min)) do*

*$0min = $0temp;*

Оператор цикла:

*loop < Оператор\_или\_блок > until < Логическое\_выражение >*

Выражение создает цикл, который выполняет указанное выражение до тех пор, пока условие не станет ложным. Условие проверяется после выполнения выражения, то есть выражение выполнится как минимум один раз.

Пример:

*$0i = 0;*

*loop {*

*$0i = $0i + 1;*

*$0result = $0result + $0i;*

*} until ($0i < 5);*

Оператор переключателя:

*select <Произвольное выражение> case (<Константа> ) <Оператор или блок> [break;] …[ case () <Оператор или блок> ] end*

Оператор *break* передает управление оператору end, который заканчивает выполнение оператора переключателя и передает управление следующему оператору.

Пример:

*long $0n;*

*$0scanf("%d", &$0n);*

*select ($0n)*

*case(1) {*

*$0do();*

*} break;*

*case(2) {*

*$0i =1;*

*} break;*

*end*

**Управляющие последовательности**

Строки и символьные константы могут содержать "управляющие последовательности".

Управляющие последовательности — это комбинации символов, состоящие из разделительного символа и не графических символов.

Управляющая последовательность состоит из знака обратного деления (\) за которым следует буква или комбинация цифр.

Управляющие последовательности обычно используются для задания каких-либо действий (например, возврат каретки или сдвиг на табуляцию на экране или на принтере) или для литерального представления неграфических символов и символов, имеющих специальное значение (например, двойные цитатные скобки - ").

Пример управляющих последовательностей:

\r – возврат каретки  
\n – перенос строки  
\t – строчная табуляция

**Знаки пунктуации и операторы**

Знаки пунктуации и специальные символы из набора символов имеют самое разное предназначение, от организации текста программы до определения задач, которые будут выполнены компилятором или скомпилированной программой.

Существуют следующие знаки пунктуации и операторы:

; точка с запятой

“ двойные кавычки

‘ одинарные кавычки

( левая круглая скобка

) правая круглая скобка

[ левая прямоугольная скобка

] правая прямоугольная скобка

{ левая фигурная скобка

} правая фигурная скобка

< меньше

> больше

<= меньше равно

>= больше равно  
!= не равно

/ знак деления  
/= поделить равно

+ плюс

+= плюс равно

\* знак умножения

\*= умножить равно

- минус

-= минус равно

== равно

++ инкремент

- - декремент

**Комментарии**

Комментарии — это последовательность символов, которая рассматривается компилятором как разделительный символ, но игнорируется им.

Комментарии могут быть везде, где допустимо появление разделительных символов.

Привет:

*//Комментарий*

**Ключевые слова**

Ключевые слова - это заранее определенные идентификаторы, которые имеют для компилятора специальное значение. Их можно использовать только так, как они определены. Имя элемента программы не может совпадать написанию с ключевым словом.

Существуют следующие ключевые слова:  
*long*

*long \_u*

*number*

*char*

*void*

*return*

*loop*

*until*

*select*

*case*

*break*

*end*

*at*

*do*

*else*

Переопределить ключевые слова нельзя.

# Вывод

В ходе лабораторной работы были описаны основные свойства языка и разобраны основные операторы рассматриваемого языка. В результате лабораторной работы была разработана тестовая программа.

# Приложение

Текст xml-файла с системой правил:

<?xml version="1.0" encoding="windows-1251"?>

<transLab>

<lexic>

<automat name='main'>

<rule groupWordsName='Addres'>

<expression>[&]</expression>

<comment>Разименовывание</comment>

</rule>

<rule groupWordsName='CharConst'>

<expression>['][][']</expression>

<comment>Символьная константа</comment>

</rule>

<rule groupWordsName='Comma'>

<expression>[,]</expression>

<comment>Запятая</comment>

</rule>

<rule groupWordsName='Comments'>

<expression>[/][/][]\*[\n]</expression>

<action>ignoreLastWord=true;</action>

<comment>Комментарии</comment>

</rule>

<rule groupWordsName='DeсadConst'>

<expression>[0-9]+</expression>

<comment>Константы по основанию 10</comment>

</rule>

<rule groupWordsName='FloatConst'>

<expression>[0-9]+[.][0-9]+</expression>

<comment>Вещественная константа</comment>

</rule>

<rule groupWordsName='Format'>

<expression>[ \t\r\n]+</expression>

<action>ignoreLastWord=true;</action>

<comment>Форматирование</comment>

</rule>

<rule groupWordsName='FourConst'>

<expression>[4][x][0-3]+</expression>

<comment>Константы по основанию 4</comment>

</rule>

<rule groupWordsName='Hooks'>

<expression>[{}()[\]]</expression>

<comment>Скобки</comment>

</rule>

<rule groupWordsName='Identificators'>

<expression>[$][0-9]+[A-Za-z]+</expression>

<comment>Идентификаторы</comment>

</rule>

<rule groupWordsName='IncludeFiles'>

<expression>[<][a-z]+[.a-z]\*[>]</expression>

<comment>Подключаемые файлы</comment>

</rule>

<rule groupWordsName='KeyWaords'>

<expression>("long""\_u"?)|("number")|("char")|("=")|("at")|("do")|("else")|("loop")|("until")|("select")|("case")|("break")|("end")|("return")|("void")|("const")</expression>

<comment>Ключевые слова</comment>

</rule>

<rule groupWordsName='Limiter'>

<expression>[;]</expression>

<comment>Ограничитель</comment>

</rule>

<rule groupWordsName='LogicOperator'>

<expression>("++")|("--")|(">")|("<")|(">=")|("<=")|("||")|("&&")|("!=")|("==")|("!")</expression>

<comment>Логические операторы</comment>

</rule>

<rule groupWordsName='Operation'>

<expression>[-+\*/]</expression>

<action>

</action>

<comment>Операции</comment>

</rule>

<rule groupWordsName='Pointer'>

<expression>[\*][$][0-9]+[A-Za-z]+</expression>

<comment>Указатель</comment>

</rule>

<rule groupWordsName='PreProcKeyWords'>

<expression>[#]("include")|("define")</expression>

<comment>Ключевые слова препроцессора</comment>

</rule>

<rule groupWordsName='SixTeenConst'>

<expression>[1][6][x][0-9A-F]+</expression>

<comment>Константы по основанию 16</comment>

</rule>

<rule groupWordsName='StringConst'>

<expression>["][]\*["]</expression>

<comment>Строковая константа</comment>

</rule>

<rule groupWordsName='Operation'>

<expression>[-+\*/%][=]?</expression>

<comment>Операции с присваиванием</comment>

</rule>

</automat>

</lexic>

</transLab>