МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ государственное БЮДЖЕТНОЕ

образовательное учреждение

высшего образования

«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Факультет автоматики и вычислительной техники

Кафедра вычислительной техники



**ОТЧЁТ**

**по лабораторной работе №2**

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

по дисциплине: «Теория формальных языков и компиляторов»

Вариант № 144211511, файл Lab2

Выполнил:Проверил:

студент гр. АВТ-918 Малявко А.А.

Ванин Константин «\_\_\_» \_\_\_\_\_\_\_\_\_ 2021 г.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(оценка, подпись)

Новосибирск

2022

# Цель работы

Изучить конечные автоматы (КА) без памяти, способы определения КА в трансляторах – графового и табличного, методы построения недетерминированного КА по системе регулярных выражений, методы эквивалентных преобразований недетерминированных КА в оптимальные полностью определенные КА – лексические акцепторы.

# Ход работы

## Результаты разработки фрагмента системы правил языка:

Видимые/редактируемые:

ident [a-z] [0-9] {1,4} [a-zA-Z] +

constdec [0-9] +

constreal [0-9] + [.] [0-9] \*

constexh ([0-9] + ([0-9] + [.] [0-9] \*)\* [e] [-] ? [0-9] +)

constchar ["] [] ["]

oper ([-+/\*])|([!=][=])|([<>][=]?)

space [ \n\r\t] + {ignoreLastWord=true;}

bracket [{}()]

comment [%][]\*[\n\r]{ignoreLastWord=true;}

delimiter [;]

keyword [a-z]+

switchword [:]

## Граф состояний и переходов сканера

Таблица 1 - Граф состояний и переходов сканера

|  |  |  |
| --- | --- | --- |
| Состояния | | | Переходы |
| (вершины графа): | | | (дуги графа): |
| **0:** | | | *EOF*->**-1**   [:]->**-2**   [a-z]->**1**   [;]->**-4**   [%]->**2**   [{}()]->**-5**   [\d32\d10\d13\d09]->**3**   [-+/\*]->**-7**   [!=]->**4**   [<>]->**5**   ["]->**6**   [0-9]->**7** |
| **1:** | | | [*other*]->**-3**   [a-z]->**8**   [0-9]->**9** |
| **2:** | | | [*other*]->**2**   [\d10\d13]->**-9** |
| **3:** | | | [*other*]->**-6**   [\d32\d10\d13\d09]->**3** |
| **4:** | | | [=]->**-7** |
| **5:** | | | [*other*]->**-7**   [=]->**-7** |
| **6:** | | | [*other*]->**10** |
| **7:** | | | [*other*]->**-8**   [0-9]->**11**   [e]->**12**   [.]->**13** |
| **8:** | | | [*other*]->**-3**   [a-z]->**8** |
| **9:** | | | [0-9]->**14**   [a-zA-Z]->**15** |
| **10:** | | | ["]->**-12** |
| **11:** | | | [*other*]->**-8**   [0-9]->**11**   [e]->**12**   [.]->**16** |
| **12:** | | | [-]->**17**   [0-9]->**18** |
| **13:** | | | [*other*]->**-10**   [0-9]->**13** |
| **14:** | | | [a-zA-Z]->**15**   [0-9]->**19** |
| **15:** | | | [*other*]->**-11**   [a-zA-Z]->**15** |
| **16:** | | | [*other*]->**-10**   [0-9]->**20**   [e]->**12** |
| **17:** | | | [0-9]->**18** |
| **18:** | | | [*other*]->**-13**   [0-9]->**18** |
| **19:** | | | [a-zA-Z]->**15**   [0-9]->**21** |
| **20:** | | | [*other*]->**-10**   [0-9]->**20**   [e]->**12**   [.]->**22** |
| **21:** | | | [a-zA-Z]->**15** |
| **22:** | | | [0-9]->**23**   [e]->**12** |
| **23:** | | | [0-9]->**23**   [e]->**12**   [.]->**22** |

## Управляющая таблица сканера

Таблица 2 – Управляющая таблица сканера

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Входные символы: | Управляющая таблица: | | | | | | | | | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** |
| [ EOF ] | **-1** | **-3** | 2 | **-6** | ***E*** | **-7** | 10 | **-8** | **-3** | ***E*** | ***E*** | **-8** | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
| [ : ] | **-2** | **-3** | 2 | **-6** | ***E*** | **-7** | 10 | **-8** | **-3** | ***E*** | ***E*** | **-8** | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
| [ a-d f-z ] | 1 | 8 | 2 | **-6** | ***E*** | **-7** | 10 | **-8** | 8 | 15 | ***E*** | **-8** | ***E*** | **-10** | 15 | 15 | **-10** | ***E*** | **-13** | 15 | **-10** | 15 | ***E*** | ***E*** |
| [ ; ] | **-4** | **-3** | 2 | **-6** | ***E*** | **-7** | 10 | **-8** | **-3** | ***E*** | ***E*** | **-8** | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
| [ % ] | 2 | **-3** | 2 | **-6** | ***E*** | **-7** | 10 | **-8** | **-3** | ***E*** | ***E*** | **-8** | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
| [ () { } ] | **-5** | **-3** | 2 | **-6** | ***E*** | **-7** | 10 | **-8** | **-3** | ***E*** | ***E*** | **-8** | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
| [  \d09  \d32 ] | 3 | **-3** | 2 | 3 | ***E*** | **-7** | 10 | **-8** | **-3** | ***E*** | ***E*** | **-8** | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
| [ \*+ / ] | **-7** | **-3** | 2 | **-6** | ***E*** | **-7** | 10 | **-8** | **-3** | ***E*** | ***E*** | **-8** | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
| [ ! ] | 4 | **-3** | 2 | **-6** | ***E*** | **-7** | 10 | **-8** | **-3** | ***E*** | ***E*** | **-8** | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
| [ < > ] | 5 | **-3** | 2 | **-6** | ***E*** | **-7** | 10 | **-8** | **-3** | ***E*** | ***E*** | **-8** | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
| [ " ] | 6 | **-3** | 2 | **-6** | ***E*** | **-7** | 10 | **-8** | **-3** | ***E*** | **-12** | **-8** | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
| [ 0-9 ] | 7 | 9 | 2 | **-6** | ***E*** | **-7** | 10 | 11 | **-3** | 14 | ***E*** | 11 | 18 | 13 | 19 | **-11** | 20 | 18 | 18 | 21 | 20 | ***E*** | 23 | 23 |
| [  \d13  \d10 ] | 3 | **-3** | **-9** | 3 | ***E*** | **-7** | 10 | **-8** | **-3** | ***E*** | ***E*** | **-8** | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
| [ = ] | 4 | **-3** | 2 | **-6** | **-7** | **-7** | 10 | **-8** | **-3** | ***E*** | ***E*** | **-8** | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
| [ e ] | 1 | 8 | 2 | **-6** | ***E*** | **-7** | 10 | 12 | 8 | 15 | ***E*** | 12 | ***E*** | **-10** | 15 | 15 | 12 | ***E*** | **-13** | 15 | 12 | 15 | 12 | 12 |
| [ . ] | ***E*** | **-3** | 2 | **-6** | ***E*** | **-7** | 10 | 13 | **-3** | ***E*** | ***E*** | 16 | ***E*** | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | 22 | ***E*** | ***E*** | 22 |
| [ A-Z ] | ***E*** | **-3** | 2 | **-6** | ***E*** | **-7** | 10 | **-8** | **-3** | 15 | ***E*** | **-8** | ***E*** | **-10** | 15 | 15 | **-10** | ***E*** | **-13** | 15 | **-10** | 15 | ***E*** | ***E*** |
| [ - ] | **-7** | **-3** | 2 | **-6** | ***E*** | **-7** | 10 | **-8** | **-3** | ***E*** | ***E*** | **-8** | 17 | **-10** | ***E*** | **-11** | **-10** | ***E*** | **-13** | ***E*** | **-10** | ***E*** | ***E*** | ***E*** |
|  | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |
| Финальные состояния и действия: | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | **N** | **Токен** | **Группа слов** |  | **Действие** | | **-1** | -> 0 | **EOF** |  | *Lexem.groupIndex=0;* | | **-2** |  | **switchword** |  |  | | **-3** |  | **keyword** |  |  | | **-4** |  | **delimiter** |  |  | | **-5** |  | **bracket** |  |  | | **-6** |  | **space** |  | {ignoreLastWord=true;} | | **-7** |  | **oper** |  |  | | **-8** |  | **constdec** |  |  | | **-9** |  | **comment** |  | {ignoreLastWord=true;} | | **-10** |  | **constreal** |  |  | | **-11** |  | **ident** |  |  | | **-12** |  | **constchar** |  |  | | **-13** |  | **constexh** |  |  | | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |

|  |
| --- |
|  |
| Обозначения в клетках управляющей таблицы: | |
| Положительное число ( 1 ) | Переход в рабочее состояние |
| Отрицательное число ( **-2** ) | Переход в финальное состояние (останов, обнаружено слово) |
| Отр. число на черном фоне (**-2**) | Возврат обрабатываемой литеры на вход автомата |
|  | и переход в финальное состояние (останов, обнаружено слово) |
| Красное ***E*** | Останов по ошибке |

## Истории работы табличного и графового лексических автоматов

Подадим на вход табличного автомата выражение put 0 to a1A;.

Таблица 3 – История работы табличного лексического автомата

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Такт** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
| Символ | p | u | t | \d32 |  | \d32 | 0 |  | 0 | \d32 |  | \d32 | t |  | t | o |
| Состояние | 0 | 1 | 8 | 8 | -3 | 0 | 3 | -6 | 0 | 7 | -8 | 0 | 3 | -6 | 0 | 1 |
| **Такт** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **27** | **28** | **29** |  |  |
| Символ | \d32 |  | \d32 | a |  | a | 1 | A | ; |  | ; |  | EOF |  |  |  |
| Состояние | 8 | -3 | 0 | 3 | -6 | 0 | 1 | 9 | 15 | -11 | 0 | -4 | 0 | -1 |  |  |

Подадим на вход графового автомата выражение put 0 to b2C;.

Таблица 4 – История работы графового лексического автомата

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Такт** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
| Символ | p | u | t | \d32 |  | \d32 | 0 |  | 0 | \d32 |  | \d32 | t |  | t | o |
| Состояние | 0 | 1 | 8 | 8 | -3 | 0 | 3 | -6 | 0 | 7 | -8 | 0 | 3 | -6 | 0 | 1 |
| **Такт** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **27** | **28** | **29** |  |  |
| Символ | \d32 |  | \d32 | a |  | b | 2 | C | ; |  | ; |  | EOF |  |  |  |
| Состояние | 8 | -3 | 0 | 3 | -6 | 0 | 1 | 9 | 15 | -11 | 0 | -4 | 0 | -1 |  |  |

## Описание учебного языка

1. Идентификаторы: <б><пЦ><пБ> - одна маленькая буква; последовательность цифр длины от 1 до 4, непустая последовательность букв. Примеры: a1928A, b0CD;

2. Константы:

целые по основанию 10 - представляет собой число любой длины, состоящее из цифр 0-9. Примеры: 0, 25;

вещественные - представляет собой число любой длины, состоящее из цифр 0-9, целая и дробная части которой разделены точкой. Примеры: 0., 0.125;

экспоненциальные - представляет собой число любой длины, состоящее из цифр 0-9, экспоненциальная часть которой начинается с буквы e. Отрицательный показатель начинается со знака минус. Примеры: 10e2, 10e-2, 0.1e-2;

символьные – представляет собой любой символ, заключенный в кавычки. Пример: “c”;

3. Знаки операций: -, +, /, \*;

4. Знаки сравнения: !=, ==, >, >=, <, <=;

5. Разделителем является знак «;»;

6. Комментарии начинаются с символа % и могут быть только однострочными;

7. Оператор присваивания задаётся следующим образом: **put** <В> **to**<И>**;**, где < В> – произвольное выражение, <И> – идентификатор, жирным шрифтом помечены ключевые слова;

8. Условный присваивания задаётся следующим образом: **if (**<ЛВ> **)** **then**<ОБ>**[ not** <ОБ> **]** где <ЛВ> – логическое выражение, <ОБ>**–** оператор или блок операторов. В квадратных скобках выделена часть конструкции, которая не является обязательной;

9. Оператор цикла задаётся следующим образом: **exec**<ОБ>**with**<И> **from** <К> **to** <К>**[ step**<K> **]** где <К> – константа;

10. Оператор переключателя задаётся следующим образом: **choice**<В>**option** <К> **:** <ОБ> **[ fin; ]** **[ option … ] … [ nooption** <ОБ> **]** **end**, где троеточие следует за конструкцией, которое можно повториться неограниченное число раз;

11. Объявление функций происходит следующим образом:  **[**<Тип>**] ( [**<АргЛист>**] )**<Блок> где <Тип>**–** ключевое слово типа, <АргЛист> – список формальных аргументов функции, <Блок> – блок операторов.

# Выводы

В данной лабораторной работе были изучены конечные автоматы (КА) без памяти, способы определения КА в трансляторах – графового и табличного, методы построения недетерминированного КА по системе регулярных выражений, методы эквивалентных преобразований недетерминированных КА в оптимальные полностью определенные КА – лексические акцепторы.

# Приложение

# Тестовая программа на языке, заданном на курсовую работу, содержащая все элементы языка

int a1A; % инициализация

put 0 to a1A;

if (a1A == 0)

then put 1 to a1A;

not put 2 \* 1 to a1A;

exec put 0.3145 to a1B with a1i 0 to 10 step 2;

choice a1A

option 1 : put 3 to a1A; fin;

option 2 : put 4 to a1A;

nooption put 0.1e-10 to a1C;

end

put int (int a1A, int a1E) start put a1A to a1E + 5; end on a1D