МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра вычислительной техники

Лабораторная работа №1

**по дисциплине «Теория формальных языков и компиляторов»**

«Лексика языков программирования. Регулярные выражения»

Вариант № 11213131

Группа: АВТ-709

Студент: Антонов К. О.

Преподаватель: Малявко А.А.

Новосибирск, 2020 г.

**Цель работы**

Освоение основных навыков работы с учебным пакетом программ автоматизации разработки трансляторов ВебТрансЛаб, изучение и освоение пользовательского интерфейса пакета и форматов файлов с исходными данными/результатами работы, изучение метаязыка регулярных выражений и технологии разработки систем правил определения лексики языков программирования.

**Порядок выполнения**

* Изучить интерфейс пакета ВебТрансЛаб: запуск, регистрация, состав основного окна, команды меню.
* Используя справку Вебтранслаба (команда меню «Помочь»), изучить структуру таблицы лексических правил (и расширения лексического анализатора) основного окна, приемы и способы формирования/редактирования ее содержимого, назначение колонок «имя автомата», «имя группы слов», «регулярное выражение», «действие» и «примечание».
* Освоить:

-  открытие системы лексических и/или синтаксических правил;

-  редактирование лексических правил;

-  операции сортировки таблицы лексических правил, добавления пустых строк, удаления, вырезания и вставки правил;

-  сохранение правил в файле;

* Изучить структуру текстового xml-файла, содержащего систему правил; освоить внешние операции (в частности - редактирование) с этими файлами.
* Изучить технологию разработки сложных систем регулярных выражений, ориентируясь на свой вариант задания на курсовую работу. Разработать и сохранить фрагмент системы регулярных выражений (пока без действий) для всех (или выбранной самостоятельно части) групп слов языка, заданного в курсовой работе.
* Разработать описание лексики заданного языка в качестве фрагмента расчетно-пояснительной записки к курсовой работе.
* Подготовить, сдать и защитить отчет к лабораторной работе.

**Вариант задания**

Согласно варианту задания на лабораторную работу, структура языка имеет следующий вид:

1. *Идентификаторы:*

|  |
| --- |
| **1** |
| <Б>\_<пбЦ>  (A\_, X\_12, W\_out…) |  |  |  |

1. *Константы:*

|  |
| --- |
| 1 |
| целые по основаниям 4, 10 и 16;  вещественные;  строковые |

1. *Объявления примитивных типов (целое, вещественное, символьное):*

|  |
| --- |
| **2** |
| **int[\_u]**  **float**  **letter** |

1. *Оператор присваивания:*

|  |
| --- |
| **1** |
| <И> **=** <В>; |

1. *Условный оператор:*

|  |
| --- |
| **3** |
| **when**<ЛВ> **then** <ОБ>**[ other**<ОБ> **]** |

1. *Оператор цикла:*

|  |
| --- |
| 1 |
| **loop** <ОБ> **until**<ЛВ> |

1. *Оператор переключателя:*

|  |
| --- |
| 3 |
| **switch**<В> **{** **by** <К> **do**<ОБ> **[ off; ]** …**[ any**<ОБ> **]** **}** |

1. *Формат псевдокода:*

|  |
| --- |
| **1** |
| триады <Код><Оп><Оп> |

Обозначения:

[...] – необязательная часть конструкции;

… – предшествующая часть конструкции повторяется произвольное количество раз;

< > – описание сокращений:

<Б>|<б>|<бБ>|<бЦ>|<Ц>|<пБ>|<пЦ>|<пБЦ> – одна большая буква | одна маленькая буква | одна любая буква | одна маленькая буква или одна цифра одна цифра | непустая последовательность букв | непустая последовательность цифр | возможно пустая посл-ть букв и/или цифр

<И> – Идентификатор (имя переменной / объекта); <К> – Константа;

<В> – произвольное Выражение; <ЛВ> – Логическое Выражение;

<ОБ> – Оператор или Блок; <О> – одиночный оператор; <ОП> – оператор присваивания;

<Код> – поле кода операции; <М> – поле метки; <Оп> – поле наименования операнда; <Р> – поле наименования результата

**Описание лексики заданного языка**

Лексика разрабатываемого языка построена следующим образом:

* Идентификаторы начинаются с большой буквы латинского алфавита, далее идет нижнее подчеркивание (\_), затем может идти последовательность из строчных букв или/и (арабских) цифр;
* Целые константы, имеющие четверичную и шестнадцатеричную систему счисления, начинаются с соответствующего значения их системы, и через знак «x» (лат.) указываются нужные числа. Например, 4x21, 16x07C;
* Шестнадцатеричные константы в своем алфавите кроме чисел от 0 до 9 имеют латинские буквы от A до F;
* Десятичные числа записываются без объявления системы счисления;
* Целая и дробная часть вещественной константы записываются через точку «.». При этом цепочка может быть записана с пустой дробной частью или с пустой целой частью, но обязательно с «.»;
* Строковые константы представляют собой любую последовательность, записанную между одиночных кавычек «’ ‘».

**Примеры правильных слов на заданном языке**

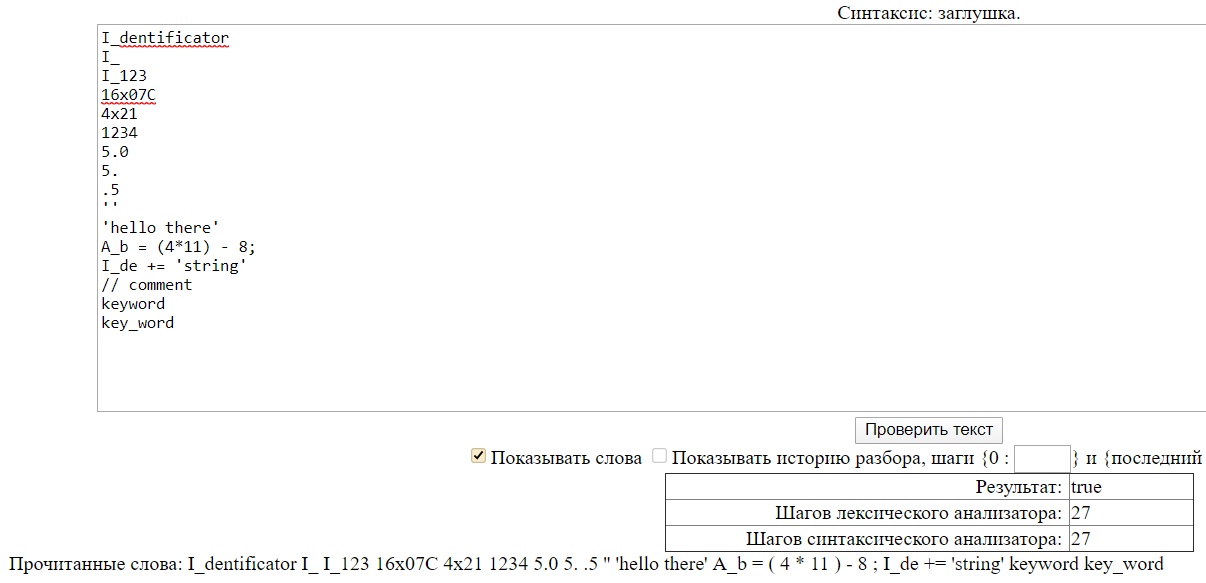


Рис. 1. Некоторые примеры слов

**Правила лексики**

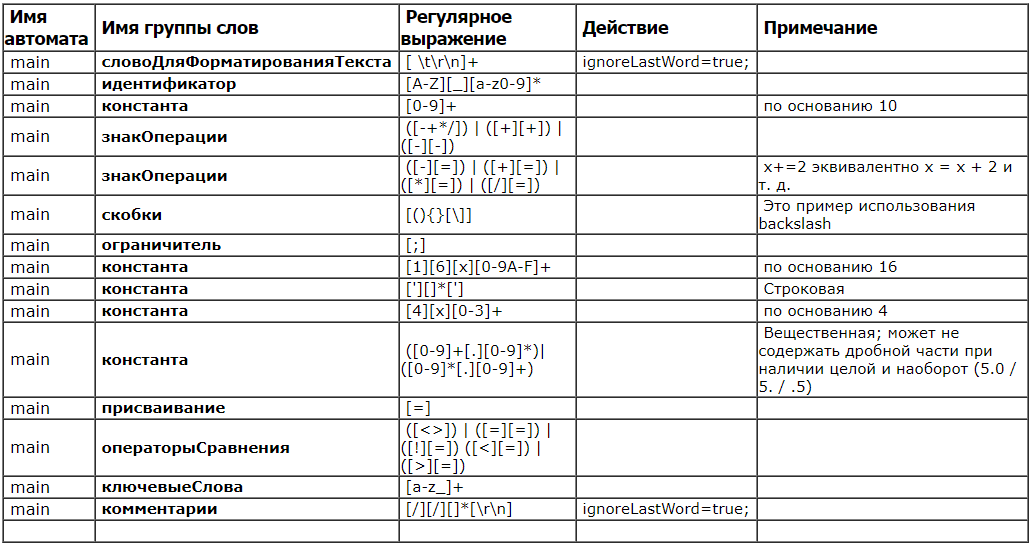


Рис. 2. Правила лексики языка

**Выводы**

В результате выполнения лабораторной работы был частично освоен учебный пакет программ автоматизации разработки трансляторов ВебТрансЛаб. В соответствии с индивидуальном вариантом курсовой работы по данной дисциплине, были разработаны правила лексики

**Приложение**

Текст xml-файла с системой правил

1. **<?xml** version="1.0" encoding="windows-1251"**?>**
2. **<transLab>**
3. **<lexic>**
4. **<automat** name='main'**>**
5. **<rule** groupWordsName='словоДляФорматированияТекста'**>**
6. **<expression>**[ \t\r\n]+**</expression>**
7. **<action>**ignoreLastWord=true;**</action>**
8. **</rule>**
9. **<rule** groupWordsName='идентификатор'**>**
10. **<expression>**[А-Я][\_][а-я0-9]\***</expression>**
11. **</rule>**
12. **<rule** groupWordsName='константа'**>**
13. **<expression>**[0-9]+**</expression>**
14. **<action>**


18. **</action>**
19. **<comment>**по основанию 10**</comment>**
20. **</rule>**
21. **<rule** groupWordsName='знакОперации'**>**
22. **<expression>**[-+\*/=]**</expression>**
23. **</rule>**
24. **<rule** groupWordsName='скобки'**>**
25. **<expression>**[(){}[\]**<>**]**</expression>**
26. **<comment>**Это пример использования backslash**</comment>**
27. **</rule>**
28. **<rule** groupWordsName='ограничитель'**>**
29. **<expression>**[;]**</expression>**
30. **</rule>**
31. **<rule** groupWordsName='константа'**>**
32. **<expression>**[1][6][x][0-9A-F]+**</expression>**
33. **<comment>**по основанию 16**</comment>**
34. **</rule>**
35. **<rule** groupWordsName='константа'**>**
36. **<expression>**['][]\*[']**</expression>**
37. **<comment>**Строковая**</comment>**
38. **</rule>**
39. **<rule** groupWordsName='константа'**>**
40. **<expression>**[4][x][0-3]+**</expression>**
41. **<comment>**по основанию 4**</comment>**
42. **</rule>**
43. **<rule** groupWordsName='константа'**>**
44. **<expression>**([0-9]+[.][0-9]\*)|([0-9]\*[.][0-9]+)**</expression>**
45. **<comment>**Вещественная; может не содержать дробной части при наличии целой и наоборот (5.0 / 5. / .5)**</comment>**
46. **</rule>**
47. **</automat>**
48. **</lexic>**
49. **</transLab>**