

Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт Космических и информационных технологий

институт

Кафедра «Информатика»

кафедра

**ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №1**

Преподаватель

Студент

КИ19–04–1М, 031943354

номер группы, зачетной книжки

\_\_\_\_\_  
подпись, дата

\_\_\_\_\_  
подпись, дата

А.С. Кузнецов

\_\_\_\_\_  
инициалы, фамилия

Т.В. Радионов

\_\_\_\_\_  
инициалы, фамилия

Красноярск 2020

## 1 Цель и задачи

*Цель:* изучение методов лексического анализа с их программной реализацией.

*Задачи:*

- изучение теоретического материала по организации лексического анализа языков программирования;
- составление формального описания программы лексического анализа;
- программная реализация по формальному описанию.

## 2 Формальное описание лексического анализатора

Описание представлено на таблице 1, где подразумевается, что регулярные выражения для лексического анализатора каждого токена расположены сверху-вниз в каждой строке и при записи в строку регулярное выражение выглядело бы следующим образом:  $/(строка\ 1) \mid (строка\ 2) \mid \dots \mid (строка\ N)/gm$

Имя токена	Регулярное выражение
INT	(?<INT>[0-9])
BOOL	(?<BOOL>\btrue\b \bfalse\b)
STRING	"(?<STRING>[^\"]*)"
PRINT IF ELSE  WHILE DO DEF  RETURN  STRUCT VAR	((?<PRINT>print) (?<IF>if) (?<ELSE>else) (?<WHILE>while) (?<DO>do)  (?<DEF>def) (?<RETURN>return) (?<STRUCT>struct)  (?<VAR>[A-Za-z]+[0-9]*))
COMMENT	\s*(?<COMMENT>.*)
PLUS MINUS  STAR SLASH  SEMICOLON  COMMA EQ  LBRACE  RBRACE  LPAREN  RPAREN  LT GT EQEQ  EXCLEQ LTEQ  GTEQ BARBAR  AMPAMP EXCL	((?<PLUS>+) (?<MINUS>-) (?<STAR>*) (?<SLASH>/)  (?<SEMICOLON>) (?<COMMA>) (?<EQ>=) (?<LBRACE>{)  (?<RBRACE>} (?<LPAREN>( (?<RPAREN>) (?<LT><  (?<GT>>) (?<EQEQ>==) (?<EXCLEQ>!=) (?<LTEQ><=) (?<GTEQ>>=)  (?<BARBAR>\\ ) (?<AMPAMP>&&) (?<EXCL>!))

## 3 Тестовые примеры

Тест 1

2 + 2

INT 2

PLUS +

INT 2

```

Тест 2
def main(a, b) { return; }
DEF def
VAR main
LPAREN (
VAR a
COMMA ,
VAR b
RPAREN )
LBRACE {
RETURN return
SEMICOLON ;
RBRACE }

```

```

Тест 3
a = 2; b = "2";
VAR a
EQ =
INT 2
SEMICOLON ;
VAR b
EQ =
STRING 2
SEMICOLON ;

```

```

Тест 4
true trues ffalse false
BOOL true
VAR trues
VAR ffalse
BOOL false

```

```

Тест 5
Привет
java.lang.Exception: Ошибка токенизации: строка 0 позиция 0:
неопознанный символ

```

```

Тест 6
def (a) { &a = 5 }
java.lang.Exception: Ошибка токенизации: строка 0 позиция 10:
неопознанный оператор

```

```

Тест 7
a1 = 1b
VAR a1
EQ =
INT 1
VAR b

```

Тест 8  
// Comment комментарий  
COMMENT Comment комментарий

Тест 9  
"text" DEF  
STRING text  
VAR DEF

Тест 10  
a = "text12{"true}"  
java.lang.Exception: Ошибка токенизации: строка 1 позиция 19:  
ожидалась закрывающая кавычка