



杭州电子科技大学  
HANGZHOU DIANZI UNIVERSITY

实验题目	词法分析器程序设计
学号	16051216
姓名	李国厦
学院	卓越学院
专业	理工类实验班
指导老师	黄孝喜

二〇一九年一月一日

## 1 实验目的

设计、编制并调试一个自定义语言 C--的词法分析程序，加深对词法分析原理的理解。

## 2 实验内容

### 2.1 自定义语言 C--的词法系统

- 1)类型系统：支持 int、char、void 基本类型，分别用词法记号表示为关键字 int、char 和 void。
- 2)常量：字符常量(用单引号括起来)、字符串常量（用双引号括起来）、八/十/六进制整数常量（0 开头表示八进制，0x 开头表示十六进制）。分别用词法记号表示为 ch、str 和 num。
- 3)变量：与常量对应，使用标识符表示，词法记号表示为 id。
- 4)表达式运算符：支持加减乘除、求余、取负、自增、自减算术运算，大于、大于等于、小于、小于等于、等于、不等于关系运算，与、或、非逻辑运算，表示为词法记号：‘+’，‘-’，‘\*’，‘/’，‘%’，‘-’，‘++’，‘--’，‘>’，‘>=’，‘<’，‘<=’，‘==’，‘!=’，‘&&’，‘||’，‘!’。注意：取负运算和减法运算在词法分析器里是被看做是同一个词法记号。
- 5)语句：支持赋值语句、do-while、while、for 循环语句，if-else、switch-case 条件分之语句、函数调用、函数返回、跳转等语句。涉及的词法记号表示为赋值号‘=’，关键字 do, while, for, if, else, switch, case, default, return, break, continue。语句和函数体要求用大括号括起来，case 和 default 后面需要跟冒号，因此需要包括各种分界符作为词法记号：‘{’，‘}’，‘;’，‘:’，‘(’，‘)’，‘,’。

所有的词法单元总结如下表所示：

表 1 C--语言的词法构成

类别	记号	含义定义	类别	记号	含义定义
标识符	id	同 C 语言标识符	运算符	div	/
常量	num	数字		mod	%
	ch	字符		inc	++
	str	字符串		dec	--
关键字	kw_int	int		not	!
	kw_char	char		and	&&
	kw_void	void		or	
	kw_if	if		assign	=
	kw_else	else		gt	>
	kw_switch	switch		ge	>=
	kw_case	case		lt	<
	kw_default	default		le	<=
	kw_while	while		equ	==
	kw_do	do		nequ	!=
	kw_for	for		comma	,
			分界符		

运算符	kw_break	break		colon	:
	kw_continue	continue		simcon	;
	kw_return	return		lparen	(
	add	+		rparen	)
	sub	-		lbrac	{
	mul	*		rbrac	}

## 2.2 单词符号编码

表 1 只是给出了建议分类，你可以根据自己程序的需要来给各种类别单词编码（可以给每个词法记号一个编码，也可以给一个类别里所有符号都编同一个编码）

## 2.3 词法分析程序的功能：

输入：所给文法的源程序字符串。

输出：二元组（编码，值）构成的序列。（能清楚表示识别出来的单词和它的类别）

# 3 实验过程

## 3.1 程序框图及算法流程

本程序的程序框图及算法流程如图 1 所示。

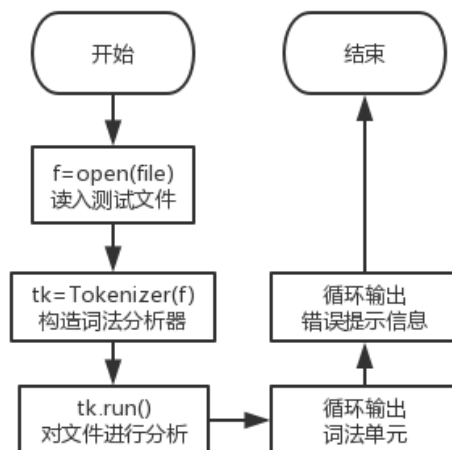


图 1 程序框图及算法流程

## 3.2 自动机

本程序构造的自动机如图 2 所示。

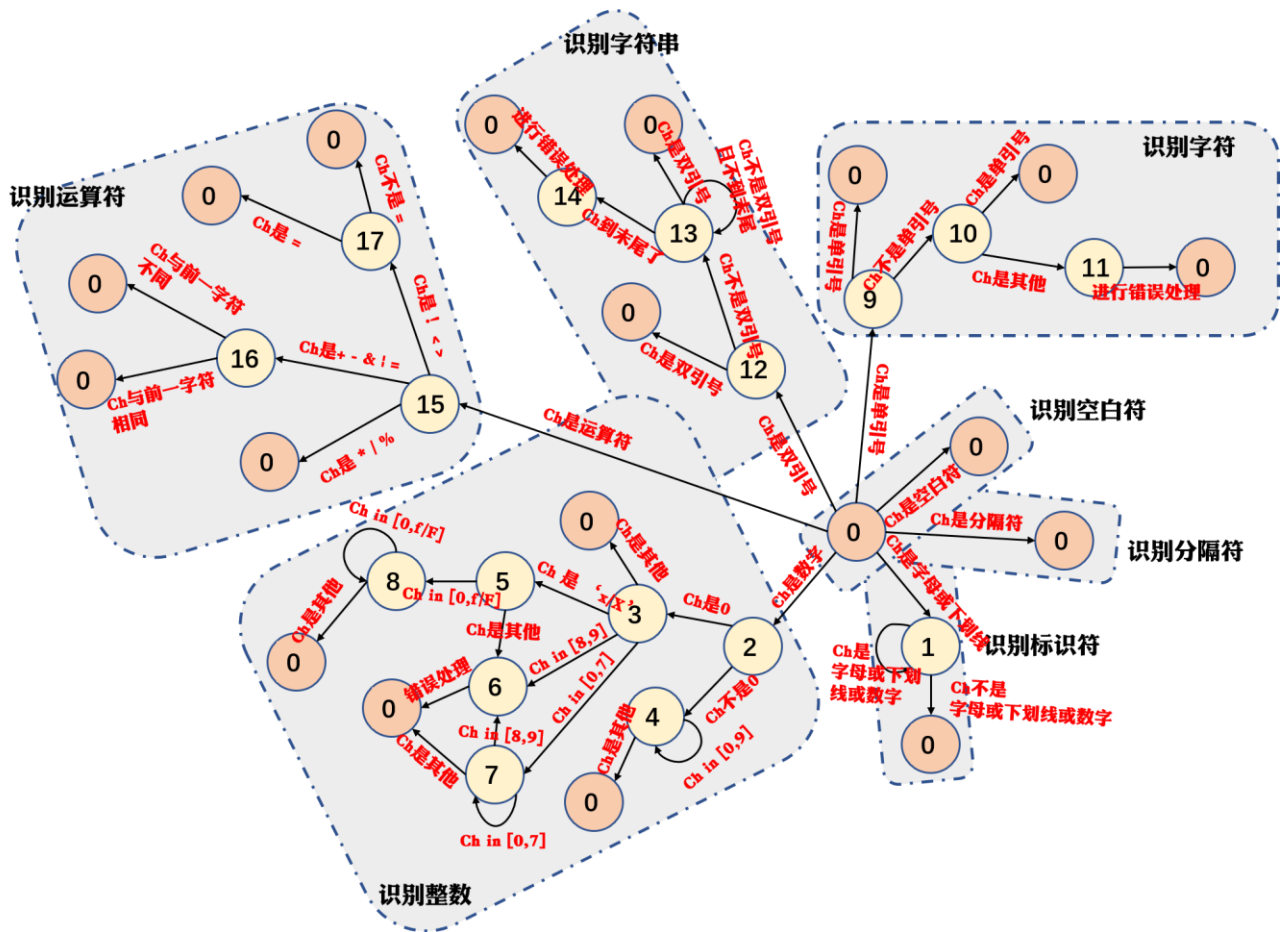
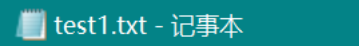


图2 自动机(0 是起始状态和终结状态)

### 3.3 实验结果

测试文件一：

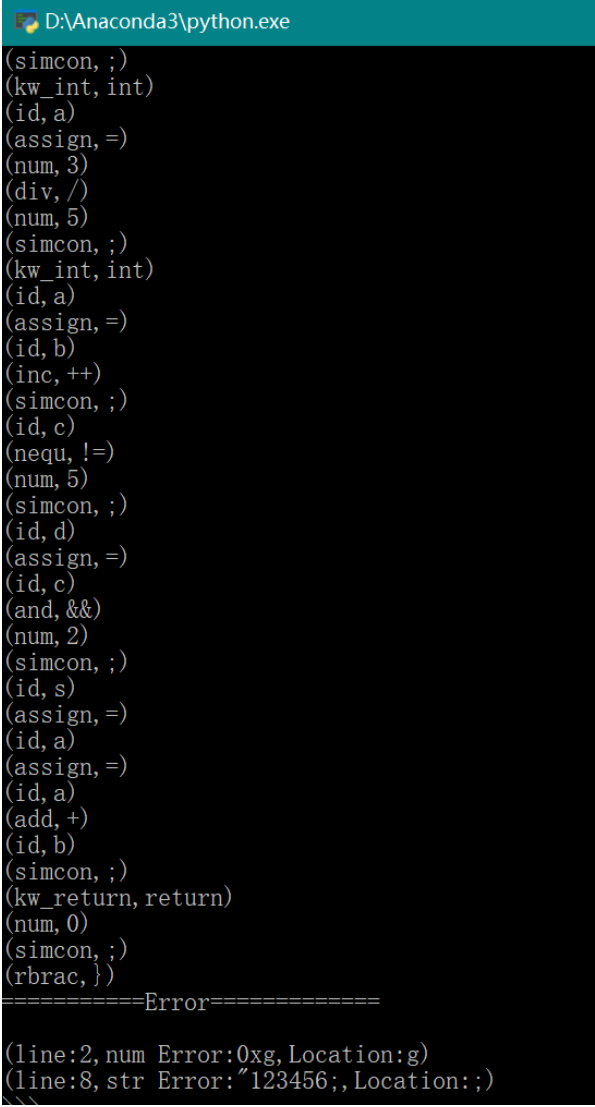
测试样例文件名	test1.txt
测试样例内容	<div>  <pre> 文件(F) 编辑(E) 格式(O) 查看(V) void main {     int a = 0x9;     int b = 6;     int a = 3/5;     int a = b++;      a = a + b;     return 0; } </pre> </div>
测试说明	没有错误的代码

实际程序运行结果

```
D:\Anaconda3\python.exe
=====Token=====
(kw_void,void)
(id,main)
(lbrac,{)
(kw_int,int)
(id,a)
(assign,=)
(int,0x9)
(simcon,;)
(kw_int,int)
(id,b)
(assign,=)
(num,6)
(simcon,;)
(kw_int,int)
(id,a)
(assign,=)
(num,3)
(div,/)
(num,5)
(simcon,;)
(kw_int,int)
(id,a)
(assign,=)
(id,b)
(inc,++)
(simcon,;)
(id,a)
(assign,=)
(id,a)
(add,+)
(id,b)
(simcon,;)
(kw_return,return)
(num,0)
(simcon,;)
(rbrac,})
>>> █
```

测试文件二：

测试样例文件名	test2.txt
测试样例内容	<pre>test2.txt - 记事本 文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)  void main {     int a = 0xg;     int2 b = 6;     int a = 3/5;     int a = b++;     c != 5;     d = c &amp;&amp; 2;      s = "123456;     a = a + b;     return 0; }</pre>

测试说明	错误的代码
实际程序运行结果	 <pre>D:\Anaconda3\python.exe (simcon,;) (kw_int,int) (id,a) (assign,=) (num,3) (div,/) (num,5) (simcon,;) (kw_int,int) (id,a) (assign,=) (id,b) (inc,++) (simcon,;) (id,c) (nequ,!=) (num,5) (simcon,;) (id,d) (assign,=) (id,c) (and,&amp;&amp;) (num,2) (simcon,;) (id,s) (assign,=) (id,a) (assign,=) (id,a) (add,+) (id,b) (simcon,;) (kw_return,return) (num,0) (simcon,;) (rbrac,}) =====Error===== (line:2,num Error:0xg,Location:g) (line:8,str Error:"123456;",Location:;) ""</pre>

## 4 实验总结

本次实验的收获主要在以下几个方面：1.完整考虑了词法分析器的需求，手动构造自动机 2.使用 if-else 实现状态转移，最终实现自动机 3.考虑了错误信息的提示。

本次实验的难点在于如何表示自动机，以及充分考虑题目需求，完整地手动构造状态图。通过本次实验，增加了对词法分析过程的理解。