# C0文法

#### 1.1单词：token

关键字（保留字，不能用作标识符）：

|  |  |  |  |
| --- | --- | --- | --- |
| int | void | main | if |
| while | return | scanf | printf |
| else | const | char | For |
|  |  |  |  |
|  |  |  |  |

特殊符号

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| + | { | ; | < | ! |
| - | } | [ | = | != |
| \* | ( | ] | >= | ， |
| / | ) | > | <= |  |

|  |  |
| --- | --- |
| 字符常量 | ＜字符＞ ::= '＜加法运算符＞'｜'＜乘法运算符＞'｜'＜字母＞'｜'＜数字＞' |
| 无符号整数 | ::= ＜非零数字＞｛＜数字＞｝ |
| 字符串常量 | ＜字符串＞ ::= "｛十进制编码为32,33,35-126的ASCII字符｝" |
| 标识符 | (英文字母|下划线)(字母|下划线|数字)\* |

词法分析器的感悟：

C语言的好多知识又复习了一遍，尤其是switch语句。

Main 函数

1.＜加法运算符＞ ::= +｜-

定义加减法运算符

例子：3+4 a+b 3+a

2.＜乘法运算符＞ ::= \*｜/

定义乘除法运算符

例子：1/2 a/2 a/b

3.＜关系运算符＞ ::= <｜<=｜>｜>=｜!=｜==

定义关系运算符

例子：1<2 a>=3 b!=8

4.＜字母＞ ::= ＿｜a｜．．．｜z｜A｜．．．｜Z

定义字母可以为大小写字母或者下划线

例子：\_ a B

5.＜数字＞ ::= ０｜＜非零数字＞

定义数字为0~9的单位数字字符

例子： 0 1 8

6.＜非零数字＞ ::= １｜．．．｜９

定义非零数字为1~9的单位数字字符

例子： 1 8

7.＜字符＞ ::= '＜加法运算符＞'｜'＜乘法运算符＞'｜'＜字母＞'｜'＜数字＞'

定义字符为以单引号开始和结束，其间有且仅有一个字符。

例子： '+' '/' 'a' '8'

8.＜字符串＞ ::= "｛十进制编码为32,33,35-126的ASCII字符｝"

定义字符串可以为0个，1个，或多个十进制编码为32,33,35-126的ASCII字符

需要注意的是不包括双引号,但是包括空格

字符串的第一个和最后一个字符为双引号

例子："hello, world~"

9.＜程序＞ ::= ［＜常量说明＞］［＜变量说明＞］{＜有返回值函数定义＞|＜无返回值函数定义＞}＜主函数＞

定义程序。

限定程序的子成分顺序，必须为常量说明在前，然后是变量说明，以上两者均可没有。

然后是有返回或者无返回的函数定义若干（可以是0个），然后才是主函数

例子: const int a = 3;

int b = 4;

void f(int a);

int f2();

void main(){...}

10.＜常量说明＞ ::= const＜常量定义＞;{ const＜常量定义＞;}

定义常量说明可以是一个或者多个const<常量定义>；这样的格式

例子： const int a = 2;

const char c = 'b';

11.＜常量定义＞ ::= int＜标识符＞＝＜整数＞{,＜标识符＞＝＜整数＞} | char＜标识符＞＝＜字符＞{,＜标识符＞＝＜字符＞}

定义常量定义可以是整型常量或者字符型常量

整型常量的定义格式是int后跟一个或多个＜标识符＞＝＜整数＞的形式，之间用逗号隔开

字符型常量同理

例子： int a = 2;

char b = 'b';

12.＜无符号整数＞ ::= ＜非零数字＞｛＜数字＞｝

定义无符号整数的开头必为非零数字，其后可跟若干个数字

该限定表明以0开头的数字串不属于无符号整数

0就不是无符号整数

例子： 1 99

13.＜整数＞ ::= ［＋｜－］＜无符号整数＞｜０

定义整数是以可省略的正负号开头，后跟无符号整数或者0的字符串

例子： 0 1 3

14.＜标识符＞ ::= ＜字母＞｛＜字母＞｜＜数字＞｝

定义标志符是必须由字母为开头，后跟0到多个字母或者数字的字符串

15.＜声明头部＞ ::= int＜标识符＞ | char＜标识符＞

定义声明头部

16.＜变量说明＞ ::= ＜变量定义＞;{＜变量定义＞;}

定义变量说明为一个或多个变量定义和分号组成的字符串

17.＜变量定义＞ ::= ＜类型标识符＞(＜标识符＞|＜标识符＞‘[’＜无符号整数＞‘]’){,(＜标识符＞|＜标识符＞‘[’＜无符号整数＞‘]’ )}

定义变量定义，变量可以为一个标志符或者由标志符为起始的数组形式。

例子： int a2bc\_d

int word[10]

但是不能如下定义：

int 34dfn

int word[-1]

18.＜类型标识符＞ ::= int | char

定义类型标识符为整数和字符这两种

19.＜有返回值函数定义＞ ::= ＜声明头部＞‘(’＜参数＞‘)’ ‘{’＜复合语句＞‘}’

定义有返回值函数。必须包含头声明头部，参数和复合语句以及必要的括号

例子： int f(char a){

...

...

}

char f(char a){

...

...

}

20.＜无返回值函数定义＞ ::= void＜标识符＞‘(’＜参数＞‘)’ ‘{’＜复合语句＞‘}’

没啥好说的。

例子： void f(char a){

...

...

}

21.＜复合语句＞ ::= ［＜常量说明＞］［＜变量说明＞］＜语句列＞

定义复合语句是现场两再变量再语句的顺序。

例子： const int a = 2;

char b = '4';

...

22.＜参数＞ ::= ＜参数表＞

。。。

23.＜参数表＞ ::= ＜类型标识符＞＜标识符＞{,＜类型标识符＞＜标识符＞}| ＜空＞

参数表由若干个类型标识符和标志符的集合组成，其间用逗号隔开。可以为空。

类型标识符和标志符的顺序不能反。

需要注意的事不能有数组形式

例如： int a, char b, int c,

24.＜主函数＞ ::= void main ‘(’ ‘)’ ‘{’＜复合语句＞‘}’

例子： void main() {

...

}

25.＜表达式＞ ::= ［＋｜－］＜项＞{＜加法运算符＞＜项＞}

表达式是由可省略的+-开头的若干个由加法运算符连接的项的字符串

例如： 3\*4+2/6-4\*a

26.＜项＞ ::= ＜因子＞{＜乘法运算符＞＜因子＞}

项是由乘法运算符连接的一个或多个因子

例如： 4\*b\*a

13\*6\*num[15]

27.＜因子＞ ::= ＜标识符＞｜＜标识符＞‘[’＜表达式＞‘]’｜＜整数＞|＜字符＞｜＜有返回值函数调用语句＞|‘(’＜表达式＞‘)’

因子是由标志符或标志符后跟方括号括起来的表达式或整数或者字符或者有返回值得函数调用语句或以圆括号括起来的表达式

例如： num[15]

num[a]

a

4

(3\*13+5\*a-2/f)

28.＜语句＞ ::= ＜条件语句＞｜＜循环语句＞｜‘{’＜语句列＞‘}’｜＜有返回值函数调用语句＞;

|＜无返回值函数调用语句＞;｜＜赋值语句＞;｜＜读语句＞;｜＜写语句＞;｜＜空＞;｜＜返回语句＞;

29.＜赋值语句＞ ::= ＜标识符＞＝＜表达式＞|＜标识符＞‘[’＜表达式＞‘]’=＜表达式＞

例如： a[3] = 4

b = 5

30.＜条件语句＞ ::= if ‘(’＜条件＞‘)’＜语句＞［else＜语句＞］

例如： if(a > b){

a = 1;

b = 2;

}

31.＜条件＞ ::= ＜表达式＞＜关系运算符＞＜表达式＞｜＜表达式＞

表达式为0条件为假，否则为真，值得注意的是一个条件中只有一个关系运算符，不支持嵌套

例如：

1&2 a&b a 0 1

32.＜循环语句＞ ::= while ‘(’＜条件＞‘)’＜语句＞| for‘(’＜标识符＞＝＜表达式＞;＜条件＞;＜标识符＞＝＜标识符＞(+|-)＜步长＞‘)’＜语句＞

循环语句具有如下形式：

while(...){

...

}

for(...;...;...){

...

}

其中大括号有时可省略

for的括号中各类型元素不能省略也不能乱位。

33.＜步长＞ ::= ＜非零数字＞｛＜数字＞}

步长为以非零数字开始数字串

形式和无符号整数类似

34.＜有返回值函数调用语句＞ ::= ＜标识符＞‘(’＜值参数表＞‘)’

例如： f(2, 3)

35.＜无返回值函数调用语句＞ ::= ＜标识符＞‘(’＜值参数表＞‘)’

同上

36.＜值参数表＞ ::= ＜表达式＞{,＜表达式＞}｜＜空＞

值参数表是若干个（包括0）表达式的集合

37.＜语句列＞ ::= ｛＜语句＞｝

语句列是若干（包括0）个连续语句的集合

38.＜读语句＞ ::= scanf ‘(’＜标识符＞{,＜标识符＞}‘)’

定义读语句是以scanf起始，后接圆括号阔起来的一个或多个以逗号隔开的标识符。

例子: scanf(a, b)

39.＜写语句＞ ::= printf ‘(’＜字符串＞,＜表达式＞‘)’| printf ‘(’＜字符串＞‘)’| printf ‘(’＜表达式＞‘)’

定义写语句是以printf为起始的，后接圆括号括起来的字符串或表达式或者两者都有，若两者都存在，则字符串在先，以逗号隔开。

例子： printf("%d", a)

printf(a)

printf("Hello, compilier!\n")

40.＜返回语句＞ ::= return[‘(’＜表达式＞‘)’]

返回语句是以return开始，后接可有可无的以圆括号包围的表达式

例子： return 0

语法分析阶段：

生成一颗抽象语法树，即操作符作为结点，操作对象作为子节点。

问题：

1. 节点类型怎么选择，怎么确定？
2. 一个非终结符号的产生体里有不止两个非终结符号，怎么构建一个二叉树。

《编译器设计》

B.3中讲了中间表示的实现

3、构建树的时候如何把括号去掉而且保证运算顺序的正确。、

＜程序＞ ::= ［＜常量说明＞］［＜变量说明＞］{＜有返回值函数定义＞|＜无返回值函数定义＞}＜主函数＞：

变量说明和有返回值函数定义有相同的声明头部。得多读两个token才能判断，这样会导致使用递归下降构造语法树变得很困难，若改造文法，那整棵树代表的意义和结构和我想象的有一定出入，所以我现在先将这个常量说明去掉，只允许const开头的常量定义，若一切进展顺利，可以回头考虑把这个变量定义加进去。

现在已经编译成功了词法分析和语法分析。

编程中遇到的困难：

一开始报了一些很奇怪的错误，比如我已经在头文件声明了parse，在c文件里定义了parse，但是报错说parse重定义，最后我发现：**包含头文件的顺序很重要，应该是我使用了条件编译的原因**

**现在有个问题：将parse.c 和parse.h 加入到项目中，生成的.exe，即使不执行parse，也会出现断言错误，不知道为什么**

**我知道了：测试程序写的有问题，这个文法不能返回0值。。。**