编译原理课程设计报告

实验题目：一个简单的C—词法分析器

一.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。

二.实验内容

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：‘{’，‘}’，‘;’，‘:’，‘(’，‘)’，‘,’。

2：词法分析程序的功能：

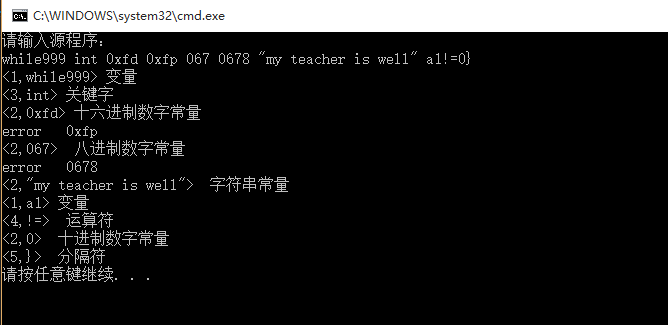
输入：所给文法的源程序字符串。

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）。

三．实验设计方案

本实验总体设计思路：逐字符识别输入字符串。遇到字符和数字时，将该字符放入str数组；当遇到其他字符时,先读入str数组内容：1.若首字母为字母，将其与关键字匹配，若匹配成功，则输出为关键字；若匹配失败，则输出为标识符id。2.若首字母为数字，则输出为数字常量；3.清空str数组等待下一次的输入。判断若输入字符为运算符，获取它的下一位输入字符，若匹配双目运算符则输出双目运算符同时将定位变量i向后移一位，否则输出该字符为单目运算符。判断若输入字符为分隔符，输出该字符为分隔符。判断若输入字符为”，在未遇到匹配的”之前，将字符串中的字符都加入str数组中，匹配到”则输出为字符串常量，再将str清空，并将定位变量向后移n位。若判断若输入字符为’，处理方式与字符串常量处理相同，输出为字符常量。

四．试验运行结果：



五．实验中遇到的问题以及改进创新

调试中发现，若输入源程序的最后一串为数字、标识符或者关键字时，无法输出。发现原因是将其输入到str数组后，由于没有遇到其他符号，便无法输出。于是在程序的最后加了一个模块用来输出以上这种情况。

实验中对数字这一块的错误输入做了判别。例如八进制中输入0~7以外的字符就是错误的，十六进制中输入0~9、A~F(a~f)以外的字符就是错误的。

验收时，发现数字0对应的是八进制常量，经过检查，发现问题。在数字识别时因为八进制是0开头且第二位不为x，于是错误的将八进制的条件设置为str[0]==0 && str[1]!=’x’，导致0输入时第一步直接进入八进制的判断当中并输出八进制数字常量。条件改进为str[0]==0 && str[1]!=’x’&& str.length()>1，即接解决了这个小问题，使之输出为十进制数字常量。

六．实验感想与体会

通过本次实验，了解了词法分析器的作用并对其有了更深的认识。本实验应该算是比较简单，但还是在初次完成后遇到了各种问题，说明自己的认识还不够到位，在前期构思的时候不够严谨，以后会更加注意。