**题目要求：**

**编写程序，能够把如下程序（prog.txt）中的词法单元都识别出来**

**int asd = 0;**

**int bc = 10;**

**while ( asd < bc)**

**{**

**if(bc - asd < 2)**

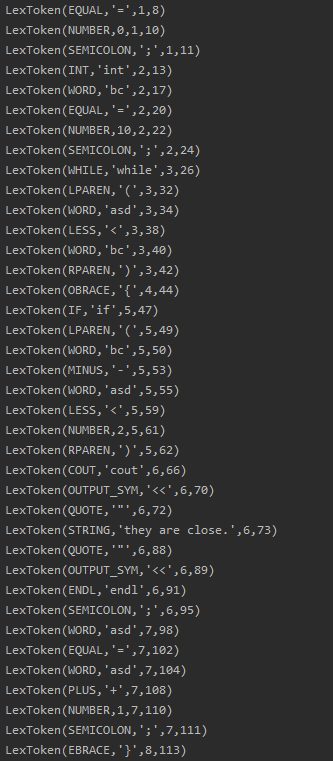
**cout<<"they are close."<<endl;**

**asd = asd + 1;**

**}**

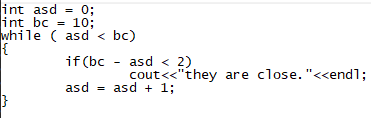
**程序说明：**

1. 打开solution.py文件，确保prog.txt和源代码在同一目录下
2. 确保已经安装了PLY库
3. 运行solution.py文件
4. 对上述txt文件中的内容词法分析，解析结果如下图：

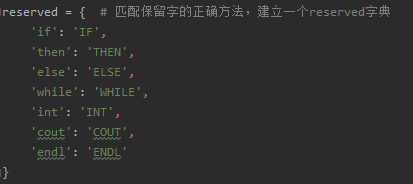


1. 对程序token定义的解释

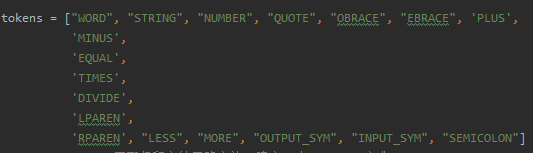
根据txt文件内容，对出现在解析中的token做了相关定义。



首先是保留字如int, while等，使用了一个保留字字典进行存储，由于容易和一般的word搞混，添加了类型检查。



其次是正常的符号、变量名等。



WORD代表正常的变量名，STRING是双引号内的内容，NUMBER是数字，QUOTE是双引号”，OBRACE是左花括号，EBRACE是右花括号，PLUS/ MINUS/EQUAL/TIMES/DIVIDE分别对应+, -, =, \*, /运算，LPAREN和RPAREN分别对应左右圆括号，LESS和MORE对应小于和大于运算，OUTPUT\_SYM对应cpp流运算中的<<，INPUT\_SYM对应>>，SEMICOLON对应分号。

1. 记处理该问题时遇到的关键难点——如何定义匹配的优先级

解决方案：对于正则字符串定义的规则来说，正则字符串越长的优先级越高；对于函数定义的规则来说，哪个放在前面哪个的优先级就高。本题中由于STRING中可能包含WORD，如果WORD的优先级高于STRING，那么they are close. 这句话就可能被匹配成三个WORD，造成问题。使STRING的优先级高于WORD即可解决。其他简单的符号匹配直接使用字符串定义就行，一般哪个长先匹配哪个。

附图：

