

**厦门大学算法实验报告**

**实验项目： 查重程序**

**专 业： 计算机科学与技术**

**年 级： 2019级**

**班 级： 1班**

**学生学号： 22920192204097**

**学生姓名： 吴雨娟**

**实验时间： 2021 年 4 月 2 日**

**一、问题描述**

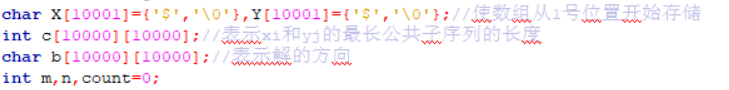
给定两个程序，如何判断他们的相似性？

**二、算法思路**

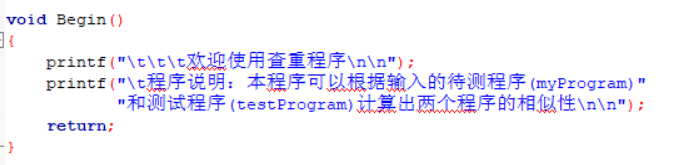
主要算法是类似最长公共子序列问题的动态规划求解，读取两个程序myProgram和testProgram，把这两个程序看成是两个序列，求出两个序列最长公共子序列的长度，再计算重复率，并输出。

**三、实现步骤**

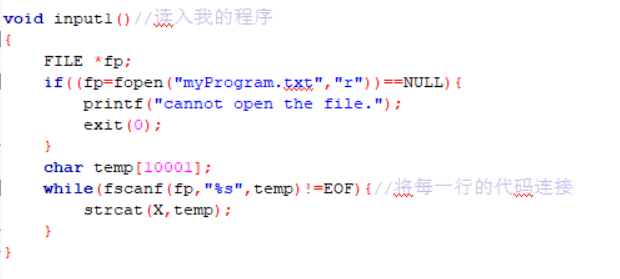
1.为了便于各个函数调用，将序列X、序列Y、记录最长公共子序列长度数组c,记录解的方向的数组b，序列X的长度m，序列Y的长度n，记录X和Y最长公共子序列长度的变量count设置为全局变量。



2.设置开始界面，包括程序标题和程序说明。

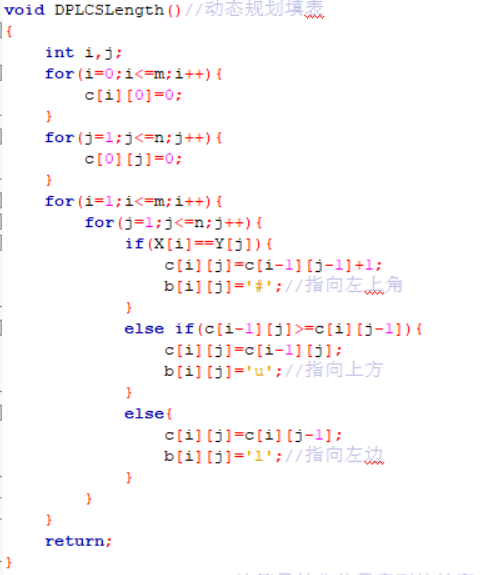


3.设置两个输入函数input1和input2,分别读入我的程序和测试程序。将程序里每一行的代码连接，形成序列X和序列Y。

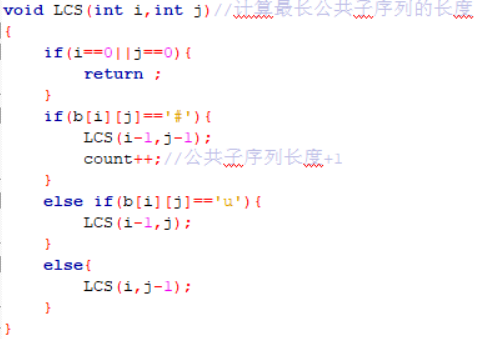




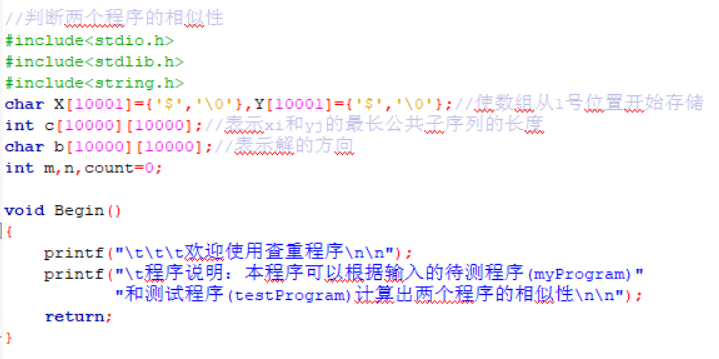
4.设置动态规划函数DPLCSLength，用自底向上方法填表c和表b。

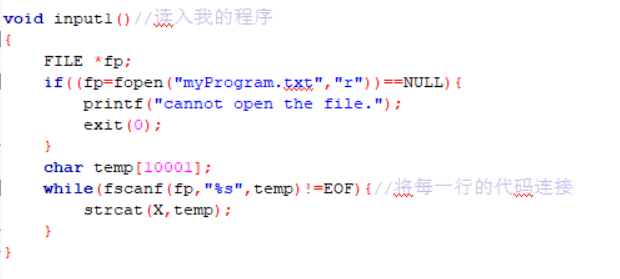


5.设置函数LCS计算最长公共子序列的长度。

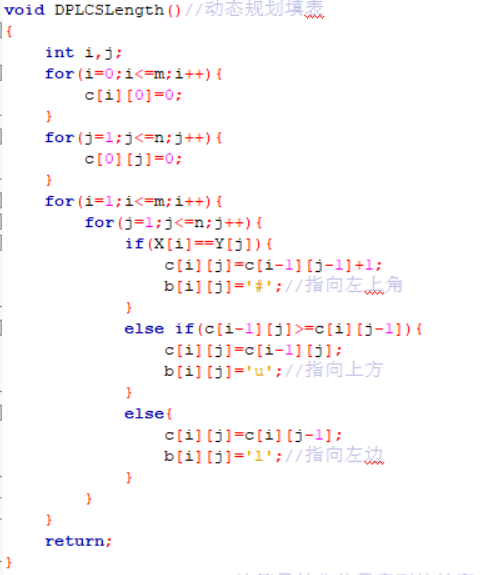


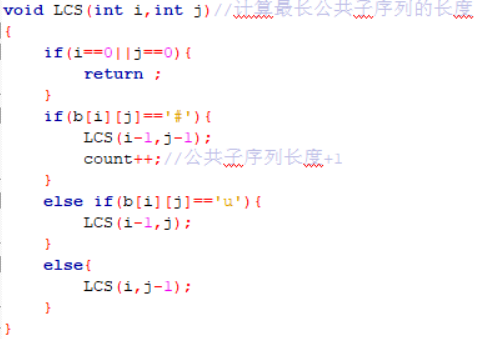
6.详细源代码如下：

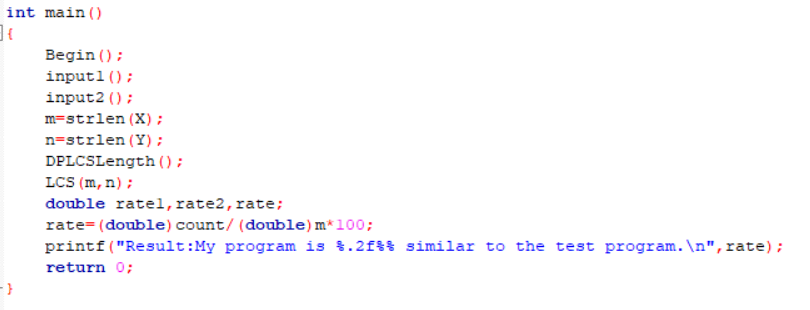








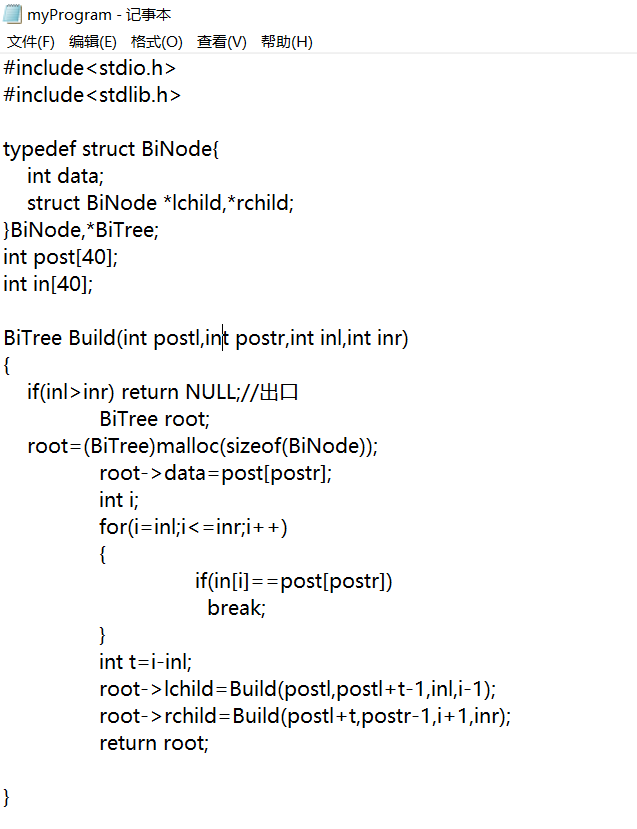


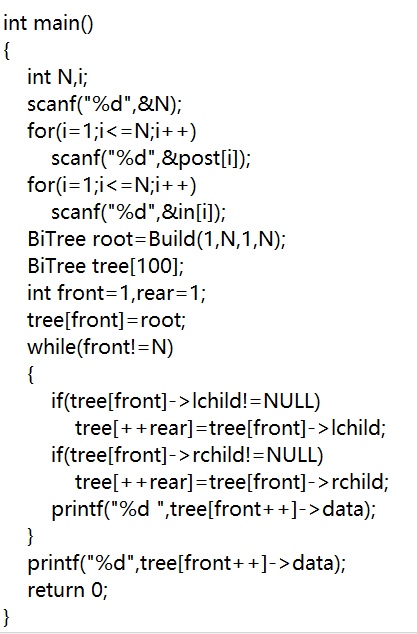


**四、运行测试**

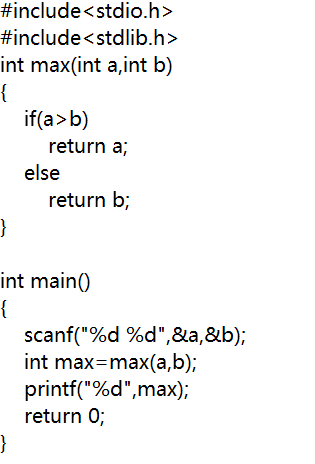
1.测试数据一（两个功能不同的程序）：

myProgram:

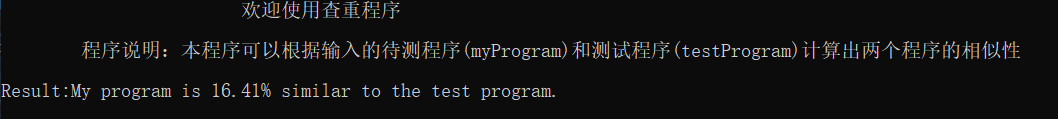




testProgram:



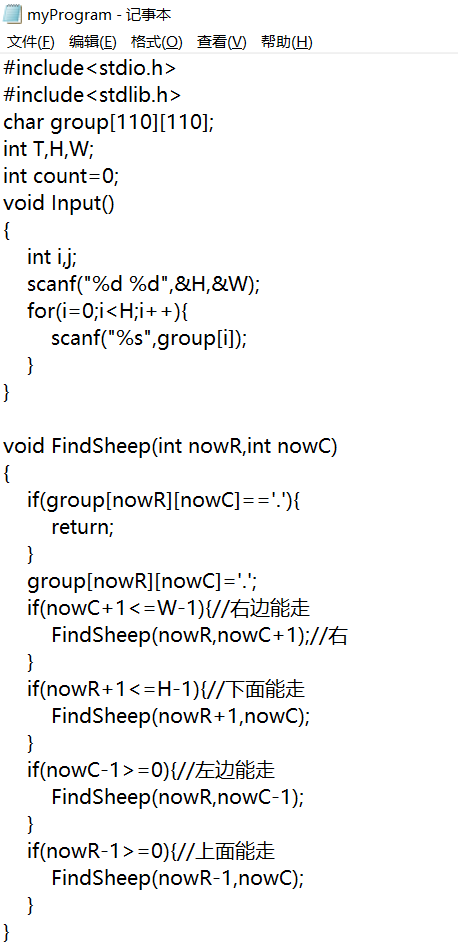
运行结果：

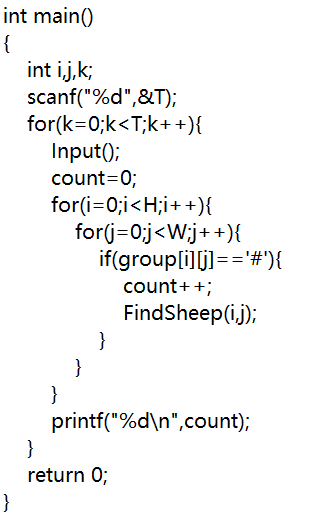


从运行结果可以看到，两个不同功能的程序重复率很低。

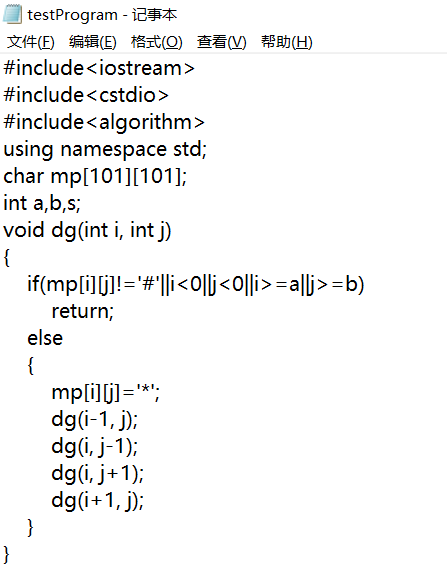
2.测试数据二（两个功能相同的程序，无抄袭）：

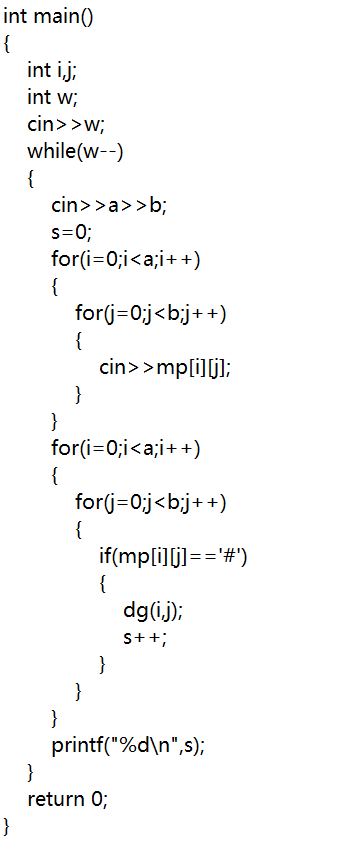
myProgram:



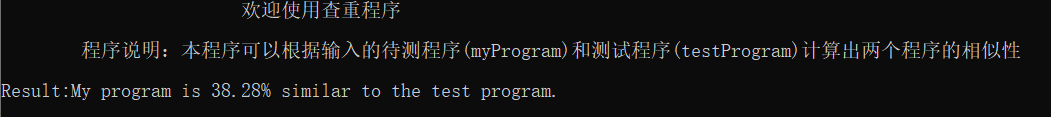


testProgram:





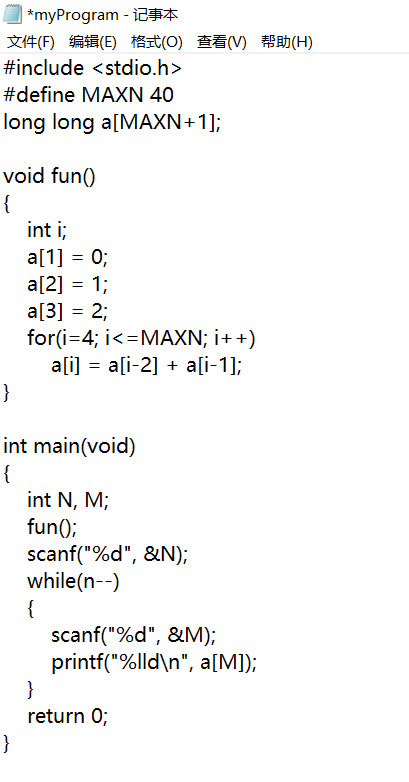
运行结果：



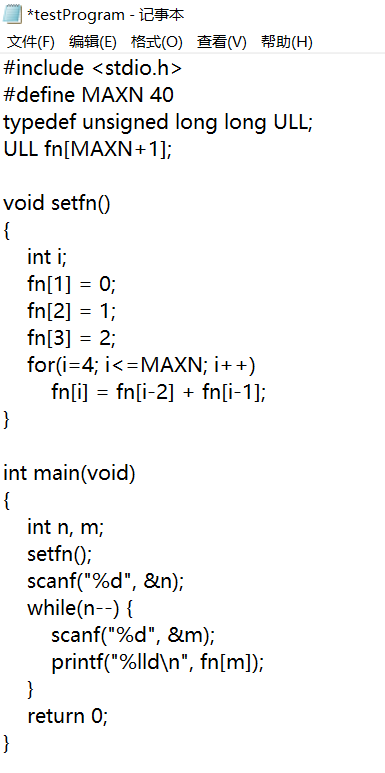
从结果可以看到，无抄袭的相同功能程序，重复率也比较低。

3.测试数据三（两个功能相同程序，有抄袭）：

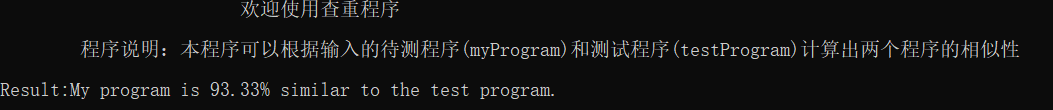
myProram:



testProgram:



运行结果：



从运行结果可以看出，抄袭的两个功能相同的程序，相似度非常高。

综上，本查重程序可以有效地区分两个程序有无抄袭，从两个程序的相似度判断抄袭的程度。

**五、结论**

本程序采用求最长公共子序列长度的方法，计算两个程序之间的相似度。该算法虽然是正确的，但是没有删除程序的一些固定标识符，导致程序相似性都略微上升了一些，不过不影响主要的判断。如果考虑算法优化，可以采用哈希表等方法。

**六、参考文献**

[1]张德富.算法设计与分析.北京:国防工业出版社,2009.