**深 圳 大 学 实 验 报 告**

**课程名称： 算法设计与分析**

**实验项目名称： 实验4 动态规划—代码查重问题**

**学院： 计算机与软件学院**

**专业： 计算机科学与技术**

**指导教师： 刘刚**

**报告人：黄志鹏 学号：2017303008 班级：计科1班**

**实验时间： 2020年5月 19日 - 2020年6月 6日**

**实验报告提交时间： 2020年6月6日**

**教务部制**

**一．实验目的**

1. 掌握动态规划算法设计思想。
2. 掌握代码查重问题的动态规划解法。
3. **实验步骤与结果**
4. **问题描述**

问题：根据给定重复比率阈值*r*与模板代码文件“*A*.cpp”*，*计算测试代码文件“*B*.cpp”的最多代码重复行数？

1. 设计动态规划算法求解文件“*A*.cpp”在第*i* 行代码与文件“*B*.cpp”在第*j*行代码的最长相同子代码模块问题，并写出求得最长相同子代码模块的递推公式。

2）对于任意给定的 *i*, *j*, 求出*S*(*i*,*j*)。根据给定重复比率阈值*r，*求出*D*(*i*,*j*)。

3）根据求出的D(i,j)，利用动态规划算法求出测试代码文件“B.cpp”的最多代码重复行数？写出求得最多行重复模板代码的递推公式。

1. **算法原理描述**
2. LCS求解最长相同子代码模块

（a）=若xm=yn, zk=xm =yn且Zk-1是Xm-1和Yn-1的一个LCS，则求Xm-1和Yn-1的LCS。

（b）若xm！=yn，求Xm-1和Yn 的LCS，Xm 和Yn-1 的LCS，由此形成两个子结构，Zk-1取其中的最大值。

（c）边界值：当求X和Y序列的子序列为0时，即i=0、j=0时，c[i，j]=0，c[i,j]定义为Xi和Yj的LCS长度，i=0~m, j=0~n。

（d）递推式：



1. 最多行重复模板代码



S(i,j)表示两行代码的相似度，LCS(A(i), B(j))为文件A在其第i行代码与文件B第j行代码LCS的长度，Li(A)为文件A在第i行代码的长度,Lj(B)为文件B在第j行代码的长度。根据矩阵S(i,j)可得到D(i,j)：



D(i,j)=1表示文件B的第j行代码和文件A的第i行代码重复，r为重复比率阈值。

3.文件预处理

-去除头文件

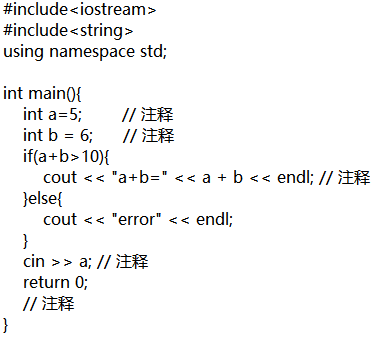
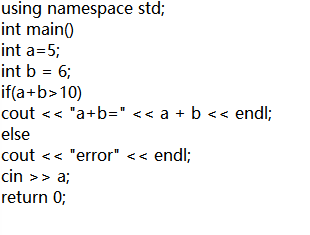
-去除注释

-去除空行

-去除多余的空格

-去除括弧

预处理前： 预处理后：



1. **算法实现的核心伪代码**

// 构造c[i][j]矩阵，表示当前的LCS长度(最优解值)

// b[i][j]为解矩阵，存放构造最优解信息

LCS\_lenght (str1, str2）

for i=0 to m; c[i,0]=0

for j=0 to n; c[0,j]=0

for i=1 to m

for j=1 to n

if (str1[i-1] == str2[j-1]) c[i][j]=c[i-1][b-1]; d[i][j]=’↖’

else if (c[i-1][j]>=c[i][]j-1) c[i][j]=c[i-1][j];d[i][j]=’↑’

else c[i][j]=c[i][j-1];d[i][j]=’←’

// 打印LCS

print\_LCS(str1, i ,j)

if i==0 || j==0; return

if b[i,j]==’↖’

print\_LCS(str1, i-1 ,j-1)

cout<<str1[i-1]

else if b[i][j]==’↑’

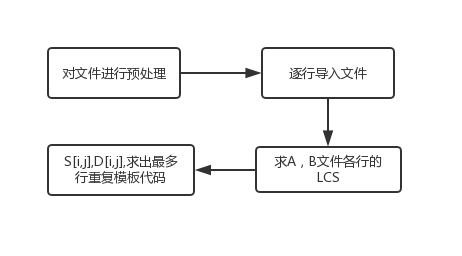
print\_LCS(str1, i-1 ,j)

else

print\_LCS(str1, i ,j-1)

1. **算法测试结果及效率分析**

(a)实验整体流程图如下所示



-其中重复率阈值设为0.7

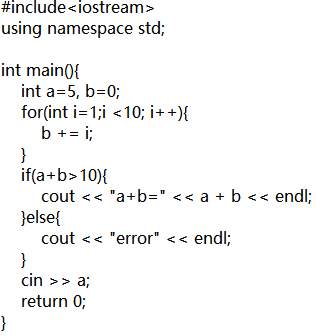
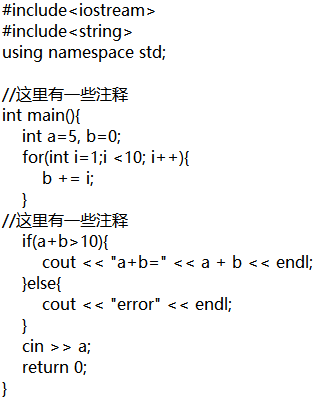
-假设代码每行不超过50个字符

（b）测试结果

对于增加注释和空行的代码查重：

|  |  |  |
| --- | --- | --- |
|  | **原代码** | **预处理代码** |
| 重复模板代码比例 | 44.4% | 90.1% |

测试样例A 测试样例B



（c）时间分析

|  |  |  |
| --- | --- | --- |
|  | 未预处理 | 预处理 |
| 耗时(ms) | 8 | 19 |

-当文件较小时，预处理耗时是为预处理的两倍。

-当文件逐渐增大时，预处理耗时所占比重将越来越小。

1. **实验心得**

-本次实验实现了利用动态规划进行代码查重的基本功能。

-本次实验还没有进行良好的优化，对替换变量名的代码查重效果不是很理想。

-总的来说，本次实验是个成功的实验，我学习并实现了动态规划求代码查重的问题，收获颇丰。

|  |
| --- |
| 指导教师批阅意见：  成绩评定：  指导教师签字：  年 月 日 |
| 备注： |

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。

2、教师批改学生实验报告时间应在学生提交实验报告时间后10日内。