

第一部分第 2 题实验指导书

2. 外排序

问题描述

应用输者树结构模拟实现外排序。

基本要求

1. 设计并实现**最小输者树**结构 ADT，ADT 中应包括初始化、返回赢者，重构等基本操作。
2. 应用最小输者树设计实现外排序，外部排序中的生成最初归并串以及 K 路归并都应用最小输者树结构实现；
3. 验证你所实现的外排序的正确性。（1）随机创建一个较长的文件作为外排序的初始数据，设置最小输者树中选手的个数，验证生成最初归并串的正确性。获得最初归并串的个数及最初归并串文件，每一最初归并串使用一个文件。（2）使用以上生成的归并串，设置归并路数，验证 K 路归并的正确性。获得 K 路归并中各趟的结果，每一趟的结果使用一个文件。
- *4. 获得外排序的访问磁盘次数，并分析其影响因素。

程序正确性的验证要求

你需要设计程序验证你所实现的外排序的正确性，本次实验我们提供了示例数据集供你验证程序的正确性。以下是关于此数据集的一些输入输出说明。

输入文件

data.in 文件为外排序的初始数据，其中第 1 行为元素个数，第 2 行开始为元素数值， properties.txt 文件中给出最小输者树的大小(树中选手的个数)和归并路数。

输出文件

output.txt 为最终排序结果， Segments 文件夹内为排序过程中产生的文件，如 Seg0-1.txt 为第 1 个初始顺串， Seg1-1.txt 为第 1 趟归并排序中产生的第一个结果文件。

数据集与实验指导书一同上传至课程群中。

此外你可以通过如下工具比较输出文件是否一致：

- Web Tools: <https://text-compare.com/>
- Windows: 在命令行中使用 FC 命令，FC file1 file2
- OSX and Linux: 在终端中使用 diff 命令，diff file1 file2

成绩

实验分数由现场验收与实验报告组成，验收时应对你所实现的代码进行讲解并当场验证你所完成的程序的正确性。以下是可能导致你成绩降低的因素：

1. 最小输者树基本功能实现不完全，缺少基本操作。
2. 对程序代码理解不清楚，对于特定的操作不能解释其实现原理。
3. 不遵守基本要求，使用其他排序算法而不是最小输者树生成最初的归并串以及完成 K 路归并。
4. 不能验证程序的正确性。程序只能接受手动输入的数据，不能使用包含大规模随机数据的文本来验证程序的正确性。

以下是获得较高成绩的几个需要考虑的方面：

1. 程序设计规范，封装良好，拥有详细的注释。
2. 对内外存读取实现了更程度的模拟可获得更高的分数。
3. 获得外排序的访问磁盘次数并进行分析。

提交详情

实验验收的截止时间为第9周实验课(4.24)，实验报告与源代码提交的截止时间为第10周实验课(5.1)晚24点。本次实验应该独立完成，请避免参考其他同学的代码，我们会对提交的代码进行查重，查重率过高会影响你的实验成绩。实验代码与实验报告请一起压缩后发送给对应助教，若未能在截止时间内提交将影响实验成绩。实验代码中无需包含测试数据，并请压缩成100M以内的压缩文件格式文件提交。

文件名命名：班级-学号-姓名-第一部分-题目序号，例：22.1-202100139999-张三-第一部分-1。

实验代码中应至少包含以下三个文件"LoserTree.h"、"LoserTree.cpp"、"main.cpp"，分别是你实现的最小输者树的头文件，最小输者树实现文件与外排序运行测试程序。其余功能的程序均可放于实验代码中一起上传。可以写一个简单的 README 来介绍各个文件的作用。