

## 实验 4 排序算法及其应用

---

### 实验项目 1: 排序算法及其应用

#### 实验题目: 内存排序算法

#### 实验内容:

排序是计算机科学中的常见任务,它将一组无序的数据元素按照某种规则重新排列,以使得数据呈现有序的状态,便于后续的查找、统计和分析等操作。当数据量较小时,将数据全部读入内存并进行排序的算法称为内存排序算法,常见的内存排序算法有:插入排序、冒泡排序、归并排序、快速排序、堆排序、基数排序等。本实验要求设计并实现上述内存排序算法并比较其运行速度。

#### 实验要求:

1. 从文本文件中将两行数据读入内存,其中第一行有一个整数  $n(n \leq 100000)$ ,表示待排序序列的长度,第二行有  $n$  个整数,用空格隔开,表示待排序序列。
2. 实现归并排序、快速排序算法,输出排序好的序列,并记录算法运行时间。
3. 实现选择排序算法或插入排序算法,并将其运行时间与归并排序、快速排序算法比较,随机生成多个适当规模的数据进行实验并绘制折线图,反映不同算法运行时间随着输入规模的变化趋势,并与理论分析结果进行比较。
4. 选做: 实现堆排序算法(基于优先级队列)和基数排序算法,并将其运行时间与上述算法进行实验比较。
5. 选做: 对于序列  $a_1, a_2, \dots, a_n$ , 若  $a_i > a_j (1 \leq i < j \leq n)$  则称  $(a_i, a_j)$  构成一个逆序对,请改写归并排序算法使其在对序列排序的同时计算原序列的逆序对数量并输出。

---

### 实验项目 2: 外存排序及其应用

#### 实验题目: 模拟外存归并排序

#### 实验内容:

许多应用场景下的输入数据条目繁多,信息量极大,无法全部放入内存,因此需要将待排序记录暂时存储在外存上,排序时将各部分分别处理。外存排序被广泛应用在数据库查询执行、大数据分析等场景,是海量数据计算领域中的一个重要技术。由于外存数据读写速度显著慢于内存数据,因此为了提升算法性能,我们需要尽量最小化算法产生的外存数据读写请求。在本实验中,我们需要用外部归并排序方法对给定文件夹中所有文件中的数据(32 位整数,数据格式保证合法)按从小到大进行排序。

#### 实验要求:

1. 使用置换-选择方法生成初始归并段,并将得到的各归并段输出为文件。使用恰当的数据结构维护每一归并段的长度信息,以便下一步使用
2. 使用上一步得到的归并段大小信息,建立最佳归并树。依据建立的最佳归并树不断做二路归并排序,最终得到归并排序结果。将结果输出为文件。记录待排序文件长度与总 I/O 次数(读/写文件的次数)。
3. 选做: 在内存使用与步骤 2 相同情况下,将二路归并排序改成  $K$  路归并排序,并重复步骤 2。记录待排序文件长度与总 I/O 次数(读/写文件的次数),并绘制折线图,与理论结果比较。
4. 选做: 重做实验,使之支持待排序文件中既有浮点数又有整数的情况。

-----  
实验说明:

1. 上传内容: (1) 源程序文件; (2) 实验测试数据和实验结果数据; (3) 实验报告 (参见“实验报告参考模板-DSA2023”)
  2. 上传格式: (1) 打包为 rar 或 zip 文件; (2) 命名规则: 学号-姓名-实验编号, 如 2022xxxxxx-李四-实验 1.rar
  3. 上传邮箱: 邮箱地址: dsafall2023@outlook.com
  4. 上传截止时间: 2023 年 12 月 9 日 (第 15 周星期六)
  5. 作业发送要求: (1) 每次使用同一个邮箱交作业; (2) 每次作业发送一封且仅一封邮件; (3) 每次实验发送一封且仅一封邮件.
-