

# 华东师范大学数据科学与工程学院上机实践报告

课程名称：算法设计与分析

年级：22 级

上机实践成绩：

指导教师：金澈清

姓名：石季凡

上机实践名称：排序算法

学号：

上机实践日期：2023.3.16

10225501403

上机实践编号：No.3

组号：1-403

## 一、目的

1. 熟悉算法设计的基本思想
2. 掌握堆排序、快速排序的基本思想，并且能够分析算法性能

## 二、内容与设计思想

1. 编程实现堆排序和快速排序。
2. 随机生成 1000、10000、100000、1000000 个数，在不同数据规模情况下，两种算法的运行时间各是多少，画图描述不同情况下的运行时间差异。

## 三、使用环境

推荐使用 C/C++ 集成编译环境。

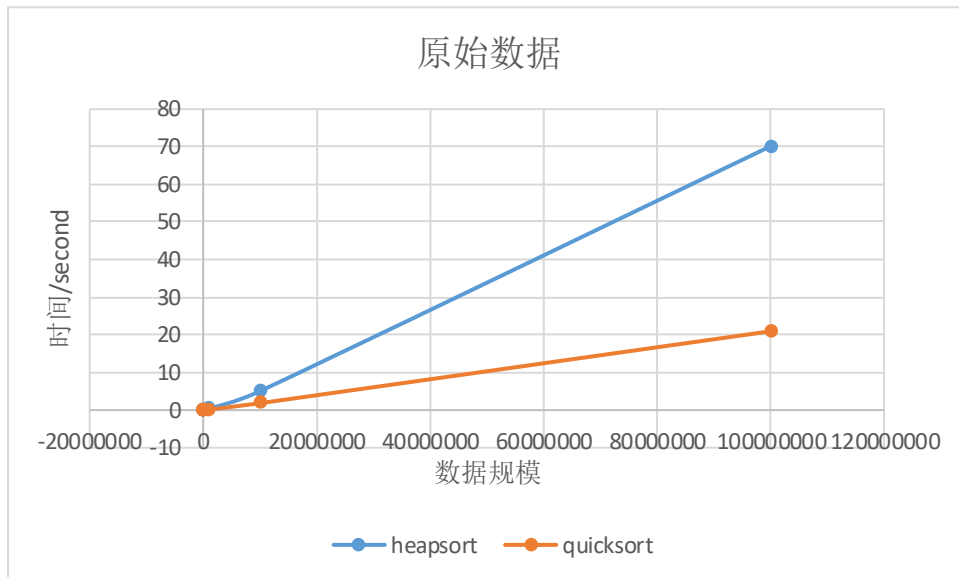
## 四、实验过程

1. 写出堆排序、快速排序算法；
2. 分别画出各个实验结果的折线图

## 五、总结

本次用 c 语言实现了堆排序与快速排序，并做了相应的优化。首先对于堆排序，实现后为了优化减少交换次数，我定义了一个变量 `largest` 用来存放最大值最后与 `parent` 交换。对于快速排序，首先为了减少循环次数我采用了前后双指针法，只需要一次循环即可完成拆分后的排序，同时为了避免固定地取中间作为分组界点会导致产生过多的顺序子数组，我采用了三者取中的方式来确定分组界点，同时由于插入排序在小规模的排序更有优势，我将分裂的小数组的长度达到一定数值时转为由插入排序进行继续排序，通过网络资料我得知，这个长度在规模较小时略大于  $\lg n$ ，在较大时约等于  $\lg n$ 。

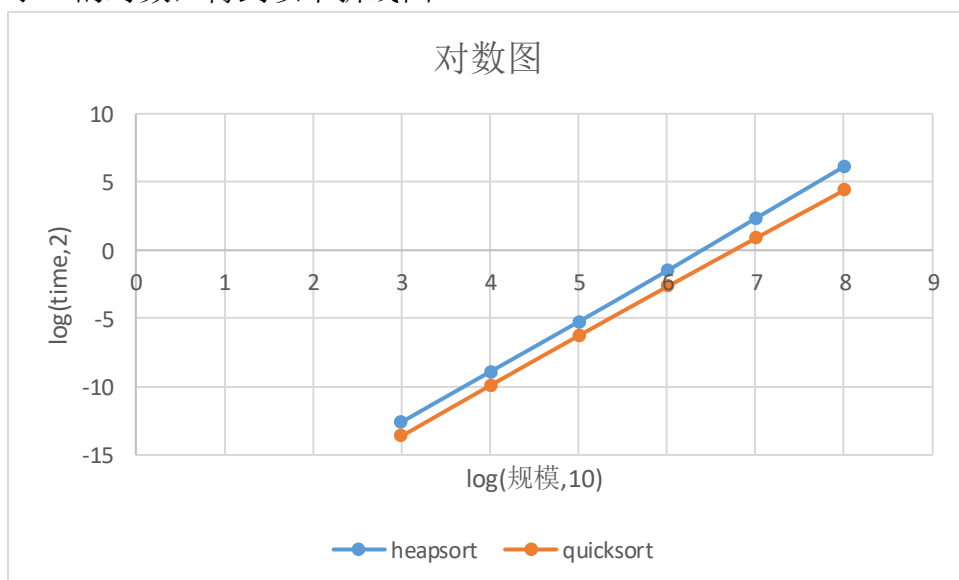
我将实验数据整合成折线图得到以下原始数据图。



具体数据如下

	heapsort	quicksort
1000	0.000162	0.000079
10000	0.002078	0.00103
100000	0.026357	0.012928
1000000	0.351297	0.156174
10000000	5.021388	1.83152
100000000	70.13141	20.95148

但是由于数据规模差度太大，我将数据规模取了关于 10 的对数，将时间取了关于 2 的对数，得到以下折线图



	heapsort	quicksort
3	-12.5917	-13.6278
4	-8.91059	-9.92314
5	-5.24567	-6.27336
6	-1.50924	-2.67877
7	2.328086	0.873041
8	6.131989	4.38898

具体数据如下

由图可见快速排序有着非常快的速度。但是在最初实现的过程中，我发现当数据规模达到百万时，运行时间会骤增，后来经检查发现是随机生成的数据范围过小，导致产生大量重复数据，最终导致快速排序的时间复杂度退化至  $n^2$ ，修改了随机数的范围后恢复正常的  $n \lg n$ 。但是同时我也发现，堆排序不论数据的类型都有着非常稳定的时间复杂度  $n \lg n$ 。

对于本次实验的反思，我对于数据生成的范围有所忽略，导致在排错时浪费了很多时间。同时，我根据学长提供的方法修改了随机数生成函数，使得原函数随机数生成的范围由原来的 0 到 32767 扩大为任意数值。