

# 数据结构与算法基础课程实验报告

## 实验 5：排序算法实验比较

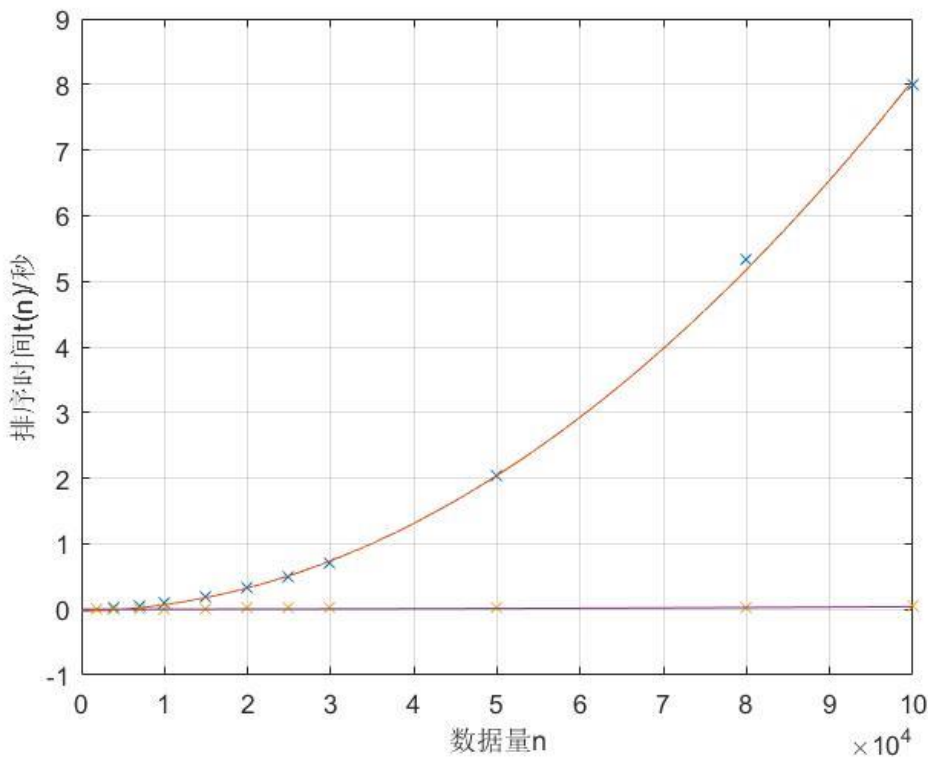
姓名	郭梓贤	院系	计算机科学与技术学院	学号	1170500112	
任课教师	臧天仪		指导教师	臧天仪		
实验地点	格物楼 207		实验时间	周六 13:30-15:30		
实验课表现	出勤、表现得分 10%		实验报告 得分 40%		实验总分	
	操作结果得分 50%					
实验目的：						
排序算法是数据处理的最基本和最重要的操作。其目的是将一组“无序”的记录序列调整为“有序”的记录序列。						
实验内容：						
实现一组经典的排序算法，通过实验数据的设计，考察不同规模和分布的数据对排序算法运行时间影响的规律，验证理论分析结果的正确性。						
实验要求：（学生对预习要求的回答）（10 分）					得分：	
直接插入排序是一种最简单的排序方法，其基本操作是将一条记录插入到已排好的有序表中，从而得到一个新的、记录数量增1的有序表。						
希尔排序是插入排序的一种又称“缩小增量排序”，是直接插入排序算法的一种更高效的改进版本，时间效率有较为明显的提升。希尔排序是非稳定排序算法。						
希尔排序是把待排序序列按下标的一定增量分组，对每组使用直接插入排序算法排序；随着增量逐渐减少，每组包含的关键词越来越多，当增量减至 1 时，整个文件恰被分成一组，算法便终止。						
实验过程中遇到的问题如何解决的？（10 分）（着重从软件调试、质量保证、结果分析方面进行阐述）					得分：	
问题 1： 插入排序比较容易实现，每次将下一个待插入的数，通过比较确定正确的位置，同时移动其他数据元素，插入适当位置即可						
希尔排序相当于通过下标取某个特定的增量，将数据分成几组分别进行插入排序，逐渐缩小增量大小，直至进行到最后一轮完整的直接插入排序。实现过程中仅需要将原来插入排序的比较过程的跨度改为希尔增量的值，并设置循环，重复多组插入排序即可。						

问题 2:

计算排序所用时间需要用到 `time.h` 头文件中 `clock` 函数，计数程序开始到当前语句所消耗的时间，在排序前后各记一次，做差即可计算出排序用时。`clock` 函数返回 `ms` 单位的数值，需要/1000 转化成秒。

每次运行程序在控制台输入不同的数据量使用 `matlab` 多项式曲线拟合，结果见下文。

本次实验的体会（结论）（10 分）	得分：
Matlab 作图结果如下，上曲线为直接插入，下曲线为希尔排序：	



可以看出两者有很明显的区别，直接插入排序的时间复杂度为  $O(n^2)$ ，希尔排序的时间复杂度与递减增量序列的选取有很大关系，在本实验中使用的是希伯特增量序列： $2^i-1$ ，希尔排序复杂度的计算是数学难题，资料给出的，使用此增量序列得到的最坏情况下的希尔排序时间复杂度为  $O(n^{1.5})$ ，例如，在数据量为 100000 时，两者时间之比应为  $100000^{0.5}$  的数量级，与实验结果符合。

数据分布情况对排序时间的影响：（30000 数据量）

	顺序	乱序	逆序
插入排序（30000 数据量）	0	0.608	1.187
希尔排序（100000 数据量）	0	0.117	0.047

可以看出插入排序在数据趋向于逆序时消耗的时间更多，意味着将进行更多次的数据移动。

而希尔排序则刚好相反，这是因为如果数据是按照完全逆序排列的，则希尔排序则能在很少的几轮之内就将数据整理成局部顺序的情形，所以大大减少了数据移动的次數

思考题：（10 分）

思考题 1：（10 分）

得分：

思考题 1：请说明实验中的排序算法的时间复杂度，并给出解释

直接插入排序的时间复杂度为  $O(n^2)$ ，直接插入排序，如果数据是非递减序列，则有最小比较次数  $n-1$ ，如果数据是非递增的，则会有最大比较次数  $(n+2)(n-1)/2$ ，记录移动次数也达到最大值  $(n+4)(n-1)/2$ ，如果记录是随机的，则是介于其之间的某个值，所以平均的比较次数和移动次数为  $n^2/4$ ，所以复杂度为  $O(n^2)$ 。

希尔排序的时间复杂度与递减增量序列的选取有很大关系，在本实验中使用的是希伯特增量序列： $2^i-1$ ，希尔排序复杂度的计算是数学难题，使用此增量序列得到的最坏情况下的希尔排序时间复杂度为  $O(n^{1.5})$ 。如果使用其他增量序列，则复杂度会有其他值。

指导教师特殊评语：

指导教师签字：

日期：