算法 LAB1 实验指南

1 比较元素选择算法

为了对分治算法的程序设计形成更为直观的认识, 以元素选择算法作为实例设计本实验, 即首先随机生成一系列不同尺寸的数据集，然后分别用排序算法、课上所讲的线性时间选择算法寻找排序在数据集 50% 位置的元素，然后计算并比较两类算法所需的时间，画出两类算法在不同数据量下的运行时间变化曲线，寻找基于分治的线性时间选择算法相对于基于排序的选择算法的性能拐点。

2 实验要求

1) 该实验要求每个人独立完成;

2) 随机生成数据集：数据集的生成不局限于采用 C/C++、Java 等语言，可以采用 Python 等生成；要求分别生成 10、100、1000、10000、100000、1000000 个元素的数据集；数据集中的元素要求随机生成，可以是正整数，最大元素值可以考虑设置为数据集的 10 倍，比如对于最大的 100 万个元素的数据集，可以将生成随机数的最大值设置为 1000 万；请务必编程时保证所生成数据的随机性;

3) 基于排序的选择算法请采用归并排序算法，比如对 100 万元素的数据集，先用归并排序算法进行排序，然后寻找位置在 50 万处的元素;

4) 无论是归并排序还是线性时间选择算法,**可以去网上找参考代码,但应局限于 C/C++、Java 等传统编程语言**，但要确保自己已读懂且代码思路与课上所讲一致（比如线性时间选择算法需要采用分成 5 个元素小组的思路，而不是随机线性时间选择）;

5) 两类算法比较要注重公平性，比如采用相同的编程语言实现（C 和 C++ 视为不同语言），采用相同的编译器及优化选项，运行在相同的硬件和系统下;

6) 数据集保存在文件中，大尺寸数据集载入非常花时间，循环迭代中的 printf 等 I/O 函数也很耗时间，统计程序的运行时间不可计入这部分时间，请采用课程开始所讲过的高精度计时工具;

7) 实验的提交物为实验报告, 实验报告要求给出自己实验的重要细节, 比如数据集的生成方法、采用的算法核心部分代码、编译及运行环境、最大数据集下运行实例的截图（程序运行完将找到的元素的数值输出出来）;

8) 最后根据得到的 6 种不同数据量下两类算法的运行时间，画出在实验所在电脑环境下运行时间随数据量变化的曲线及对比，具体图样可参考课件上比较 4 种排序算法所采取的样式。画图工具随意，python 的 matplotlib、matlab、excel 或其它工具都可以，但需要在实验报告中说明；

9) 实验报告第 8 周周一早 8 点之前提交到 canvas；务必不要抄袭，视情况对报告进行面批。