

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：21 级

上机实践成绩：

指导教师：金澈清

姓名：杨茜雅

上机实践名称：哈希

学号：

上机实践日期：

10215501435

上机实践编号：No.6

一、目的

1. 熟悉算法设计的基本思想
2. 掌握哈希算法的思想，哈希函数，哈希冲突解决等算法

二、内容与设计思想

1. 完成 OJ 上的题目。
2. 用以前写的随机数生成器生成规模为 $10^2, 10^3, 10^4, 10^5$ 的随机数据规模（数据范围 $0-10^5$ ），插入哈希表（表的大小固定），并再次生成相同规模的数据，查询哈希表中是否有该元素，并画图描述不同数据规模下的插入&查询的运行时间情况。
3. 设计方法来比较线性探查和二次探查的性能差异（请考虑数据插满的情况）。

三、使用环境

推荐使用 C/C++集成编译环境。

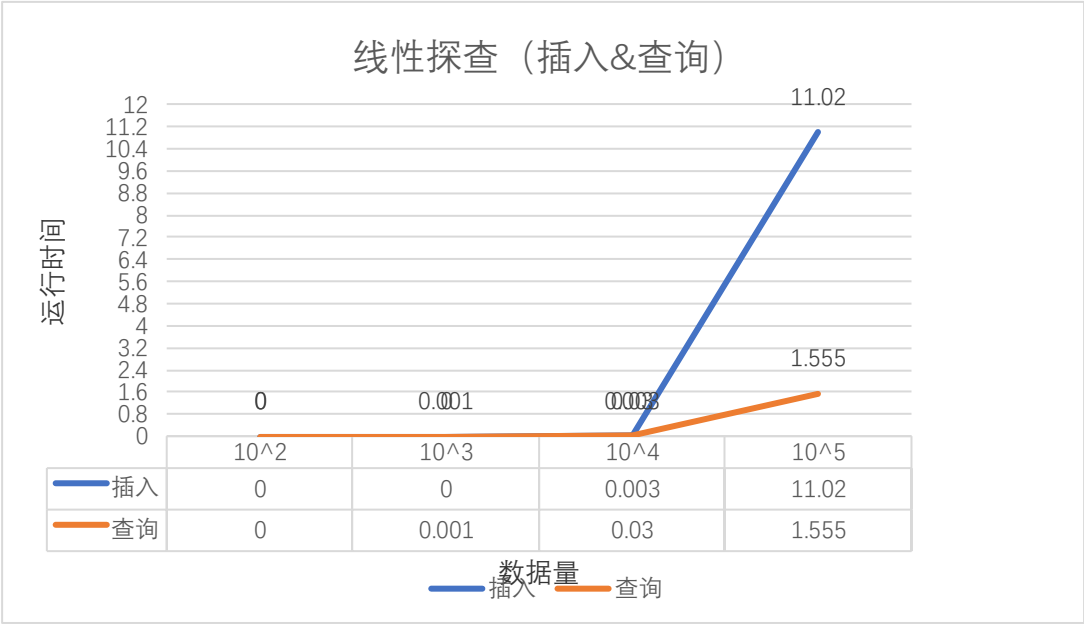
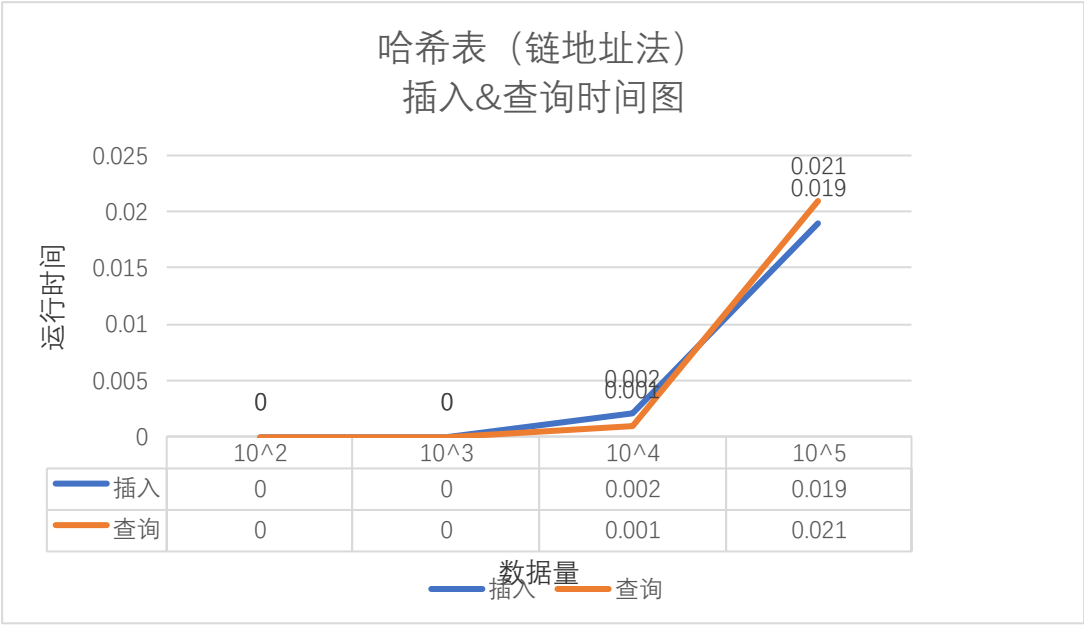
四、实验过程

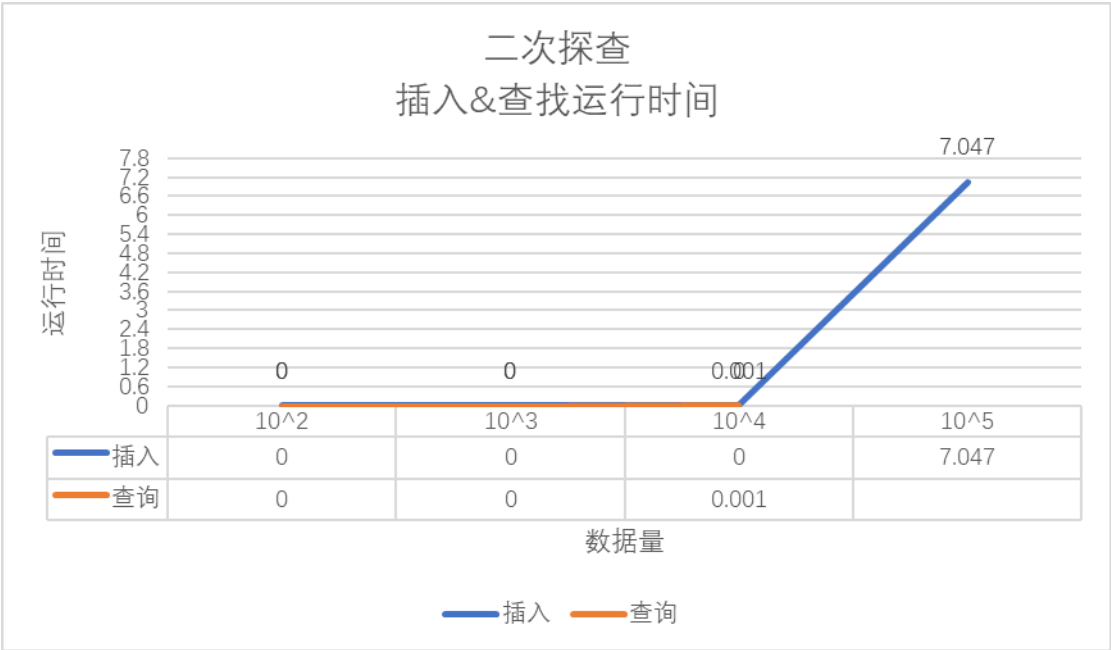
1. 编写哈希的相关代码
2. 作图展示插入&查询的运行时间折线图
3. 分析遍历数组查询，二分查询，哈希查询的差异，探讨它们各自的适用场景

二分查找：数组必须是有序的 找中间数和要查找的数字比大小 通过大小比较结果缩小查找的范围。二分查找实际上体现了解决问题的典型策略：分而治之，即将问题分为若干更小规模的部分，通过解决每一个小规模部分问题，并将结果汇总得到原问题的解，可以用递归。算法复杂度是 $O(\log n)$ 。

遍历数组查找：数组可以无序，但是遇到很不幸的情况的时候需要遍历完所有数组，耗时长，时间复杂度 $O(n)$ 。

哈希查询：适用于一些数据反应它们共同的特性比如（ $\text{mod } N$ 的结果都是相同的，这样可以直接在 $\text{array mod } n$ 链表中找，范围被缩小了），而且不需要数组有序，如果数据量很大可以开链表来减少数组的长度。删除和查找是 $O(1)$ 的复杂度，查找最大数据或者最小数据的时间复杂度是 $O(N)$ 。





（原谅我 10⁵ 查找实在是跑不出来了……）

五、总结