

大作业：小蓝鲸的财务统计任务

1. 大作业总分：5 分

2. OJ 平台截止日期：2020 年 1 月 3 日 23:59

3. 书面材料提交方式与时间：

截止日期：2020 年 1 月 5 日, 23:59, 截止日期之后不再接收提交。

提交内容：书面报告文档（Word 版以及 PDF 版本）+ 程序源码+程序使用说明

提交压缩文件及命名方式：学号_姓名_数据结构 2019 大作业.rar

提交方式：以邮件形式发送至 nju_ds_2019_1@hotmail.com (李武军老师班)

nju_ds_2019_2@hotmail.com (商琳老师班)

1. 概述

在政府工作的小蓝鲸接到了一项任务：分析蓝鲸市市民的年收入，并根据要求计算相关统计值。聪明的小蓝鲸已经想好了应该如何分析，但是数据量太大了，她需要计算机来辅助处理数据。于是，小蓝鲸向你寻求帮助，希望你能编写满足她的需求的程序。

2. 功能

2.1 统计多个收入区间内的平均收入（大作业总分的 15%）

给定待查询的收入区间，小蓝鲸需要根据系统内已录入的收入列表，统计该区间内的市民收入平均值，将其平均值取整输出。

2.2 实时统计收入的中位数（大作业总分的 15%）

每当小蓝鲸录入一个新的市民收入到系统之后，小蓝鲸需要根据录入后的收入列表，重新计算所有市民收入的中位数。

注：当 k 为奇数时，直接输出第 $(k+1)/2$ 大的收入；当 k 为偶数时，输出第 $k/2$ 和 $k/2+1$ 大的收入的平均值 ($k \geq 1$)。

2.3 实时统计当前最高的 k 个收入（大作业总分的 15%）

给定需要统计的数目 k ，每当小蓝鲸录入部分新的市民收入到系统之后，小蓝鲸需要根据系统的检查请求，实时返回此次请求时市民收入的前 k 高的收入值。

2.4 查找蓝鲸市某市民（大作业总分的 15%）

给定某个待查询的市民身份证号，小蓝鲸需要从系统内已录入的市民身份证中查找待查询的市民是否已经在系统中。

2.5 计算收入的最大断档（大作业总分的 15%）

对于系统中已录入的收入情况，小蓝鲸需要快速统计出收入的最大断档，即对收入排序后相邻两个市民收入的差值的最大值。

3. 设计要求

3.1 上述功能实现语言指定为 C++，除 String 类外，不得调用 STL 库，所有功能须自行编写类、函数进行实现。

3.2 每个功能模块，可以自行选择不同的数据结构进行存储与访问。

(1) 若满足题目给出的时间与空间复杂度（若有）且通过 OJ，该功能模块不扣分；

(2) 若通过 OJ 但时间与空间复杂度不合题意，该模块扣大作业总分的 5%；

(3) 若未通过 OJ，该模块扣大作业总分的 5%。

3.3 须按照功能模块，提交源代码，且提供 README 文件，注明代码运行方式。

3.4 每个功能模块代码注释详尽，若某一功能模块无注释，该模块扣大作业总分的 5%（扣分不超过大作业成绩的 15%）。

3.5 书面报告说明（大作业总分的 25%）

1) 描述各模块使用的数据结构、实现的思路、方法与细节；

2) 给出各个模块算法的平均时间复杂度以及平均空间复杂度；

3) 展现各个模块运行的结果并加以说明。

模块 1：统计多个收入区间内的平均收入

描述

给定 n 个市民的收入和 m 个询问区间 $[l_i, r_i]$ 。

对于每个询问区间 $[l_i, r_i]$ ，你的程序需要计算在 l_i 到 r_i 范围的收入的平均值，取整后输出。

时间复杂度

算法的时间复杂度尽量控制在 $O(m \log n)$ 。

样例数据

输入

```
5 3          // n=5 个蓝鲸市民的收入，后面 n 行，第 k 行表示第 k 个市民的收入；
              m=3 个统计区间，后面 m 行，每行两个数表示统计区间。
100000
40000
30000
60000
200000
1000 40000
30000 40000
30000 100000
```

输出

```
35000
35000
57500
```

模块 2：实时统计收入的中位数

描述

对于 n 个市民，逐一获取每个市民的年收入。当获取第 k 个年收入时，统计并输出前 k 个年收入的中位数。

注：当 k 为奇数时，直接输出第 $\frac{k+1}{2}$ 大的收入；当 k 为偶数时，输出第 $\frac{k}{2}$ 和 $\frac{k}{2} + 1$ 大的收入的平均值 ($k \geq 1$)。

时间复杂度

算法的时间复杂度尽量控制在 $O(n \log n)$ 。

样例数据

输入

```
5           // n=5 个蓝鲸市民的年收入，后面 n 行，第 k 行表示第 k 个市民的收入
100000
40000
30000
60000
200000
```

输出

```
100000
70000
40000
50000
60000
```

模块 3：实时统计当前最高的 k 个收入

描述

给定 n 个市民的年收入，以及查询参数 k ，要求实时反馈出所有收入中最高的 k 个收入。收入信息按行给出，每行表示一个市民的年收入。当该行的内容为“Check”字符串时，要求输出到上一条收入为止的所有收入中，最高的 k 个收入。如果 k 值超过了当前已经存储的收入数目，则仅需输出已有的数据。输出在一行完成。

时间复杂度

算法的时间复杂度尽量控制在 $O(n \log n)$ 。

样例数据

输入

```
10 3 // n=10 个市民，要求反馈出前 k=3 高的收入值。
10000
7000
Check
40000
2800
1500
Check
500
8000
3000
Check
2000
60000
```

输出

```
10000 7000
40000 10000 7000
40000 10000 8000
60000 40000 10000
```

模块 4：查找蓝鲸市某市民

描述

给定所有蓝鲸市民的身份证号，共 n 个。市民与身份证号一一对应。身份证号总共有 18 位，由 0~9 的数字和大写字母组成。

再给 m 个市民的身份证号，需要判断这 m 个市民是否属于蓝鲸市。

时间复杂度

算法的存储空间尽量小，每次查询的平均时间复杂度尽量控制在 $O(1)$ 。

样例数据

输入

```
5 2 // n=5 个蓝鲸市民的身份证号，后面 n 行，第 k 行表示第 k 个市民的身份证号
12345678912345678X // 2 表示 m=2，第 n+1+i 行，表示第 i 个要查询的市民身份证
98765432198765432X
123456789A12345678
987654321B98765432
987654321C98765432
123456789A12345678 // 第 1 个要查找的市民的身份证号
123456789C34545678 // 第 2 个要查找的市民的身份证号
```

输出

```
true // true 表示该市民属于蓝鲸市，false 表示该市民不属于蓝鲸市
false // true 表示该市民属于蓝鲸市，false 表示该市民不属于蓝鲸市
```

模块 5：计算收入的最大断档

描述

给定 n 个市民的年收入，要求计算收入的最大断档，即对收入排序后相邻两收入的差值的最大值。

时间复杂度

算法的时间复杂度尽量控制在 $O(n)$ 。

样例数据

输入

```
5          // n=5 个市民，后面 n 行，第 k 行的 int 型整数表示第 k 个市民的年收入
10000
40000
8000
60000
28000
```

输出

```
20000     // 一个 int 型数，代表收入的最大断档
```