

第一章 开篇

问题1

如果不缺内存，如何使用一个具有库的语言来实现一种排序算法以表示和排序集合

在C语言的实现里面写了a[1000000]，这里将占用栈的空间，C语言中栈空间的大小很有限，如果不使用动态申请堆空间的方法，有的机器可能无法运行这段程序。

```
public static void main(String[] args){

    BufferedReader reader = new BufferedReader(new FileReader("test.csv"));
    String line;
    List<Integer> list = new ArrayList<>();
    while((line = reader.readLine()) != null){
        list.add(Integer.valueOf(line));
    }
    reader.close();
    Collections.sort(list);
    for(int n : list){
        System.out.println(n);
    }
}
```

问题2

如何使用位逻辑运算(与、或、位移)来实现位向量？

一个int一共有32bit，对应 2^5 ，对于一个给定的数，首先确定其头部，可以 $i \gg 5$ 位得到，它的尾部是 $(1 \ll (i \& 0x1f))$

```

public class Bitmap {
    private final int BITSPERWORD = 32;
    private final int SHIFT = 5;
    private final int MASK = 0x1f;
    private final int MAX_NUM = 1000000;
    private int[] nums = new int[1+MAX_NUM/BITSPERWORD];

    public Bitmap(){}

    public void set(int i){
        nums[i >> SHIFT] |= 1 << (i & MASK);
    }

    public void clear(int i){
        nums[i >> SHIFT] &= ~(1 << (i & MASK));
    }

    public int test(int i){
        return nums[i >> SHIFT] & (1 << (i & MASK));
    }
}

```

问题3

// TODO

问题4

如何生成0到n-1之间的K个不同的随机整数

```

public static void main(String[] args){
    final int MAX_NUM = 1000000;
    int[] nums = new int[MAX_NUM];
    for(int i = 0; i < MAX_NUM; i++){
        nums[i] = i;
    }
    int k = 900000;
    Random random = new Random();
    for(int i = 0; i < k; i++){
        int tmp = i;
        int index = i + random.nextInt(MAX_NUM - 1 - i);
        nums[i] = index;
        nums[index] = tmp;
        System.out.println(nums[i]);
    }
}

```

问题5

如果严格限制1MB的使用空间，那么应该如何处理

这样只能使用两趟算法

问题6

如果不是每个整数至多出现一次，而是每个整数至多出现10次，应该如何处理？解决方法如何随着可用的存储空间大小的变化而变化？

不大于10的数需要至少4位来表示，因此之前的设计方案需要修改，之前使用1位来表示一个数字出现的次数现在改成4位，原来一千万个数字需要一千万位，现在需要4千万位，大约500万字节，一趟算法可用的空间只有1MB，大约是一百万字节，那么至少需要5趟算法才能完成排序

问题7

程序如何处理边界情况，某个数字出现超过一次？输入的数据大于n或者小于0，输入的数据不是数值？

// TODO

问题8

原来问题的背景是所有的免费电话的开头都是800，现在如果增加一些免费电话的开头，比如877,888，如何在1MB的空间中对这些数据进行排序，如何将免费的电话号码存储在一个集合中，可以快速的判断一个号码是否可用或者已经存在？

做法和问题6类似，不同的开头代表相同的电话号码出现了多次，如果免费电话的开头有n种，那么就需要n位向量表示各个开头出现的情况，一共总长度是 $n * 10^7$ ，大约 $125n * 10^4$ 字节，一趟算法可用的空间假如只有1MB，大约是一百万字节，那么至少需要 $1.25 * n$ (向上取整)趟算法才能完成排序

问题9

用空间换取时间的一个问题是，大量的空间初始化需要很多的时间，那么如何设计一种技术，在第一次访问向量的项时将其初始化为0。你的方案应该使用常数时间进行初始化和向量访问，使用的额外空间应该正比于向量的大小。因为该方法通过进一步增加空间来减少初始化的时间，所以仅在空间很廉价，时间很宝贵且向量很稀疏的情况下才考虑使用。

// TODO

问题10

商店的数据库使用客户的电话号码作为其检索的主关键字，那么如何组织商店的数据库，允许高效的插入和检索操作？

将电话号码的最后两位建立一个10*10的散列索引，最后两位基本是随机的，因此散列的效果

问题11

在20世纪80年代早期，洛克希德公司加利福尼亚州桑尼维尔工厂的工程师们每天都要将许多由CAD系统生成的图纸从工厂送到另外一个地方，相距40KM，但是由于地理位置的影响，用汽车送需要一个多小时，花费100美元，请给出新的数据传输的方法并进行评估。

当时没有网络传输技术，因此是否可以考虑传真技术？答案中建议使用信鸽传输微缩胶卷，也比较可靠。

问题12

载人航天的先驱们很快就意识到需要在外太空的极端环境下实现顺利书写。明见盛传NASA花费100万美元研发出了一种特殊的钢笔来解决这个问题，那么前苏联如何解决相同的问题呢？

先看看问题的难点在什么地方？我们通常使用的墨水笔是因为地球重力的缘故，写在纸上能出现字，那在外太空的环境下，没有重力或者重力很小，墨水笔就写不出字了。看上去问题就在于如何在微重力的情况下让墨水笔也能写字。

这个问题对于没什么物理学知识的人来说或许很困难，但是仔细想一下，为什么非要用墨水笔写字呢？用依靠摩擦原理的铅笔就可以解决这个问题了。在进一步想一下如果真的在极端的环境下，铅笔也没有了，那就只能用石头刻在石头上了，不由得想起了《三体3》里面保存人类历史文明的方式。