



并查集

主讲：骆明宇

复核：李梁裕 应懿 潘飞扬



什么是并查集？

并查集，在一些有 N 个元素的集合应用问题中，我们通常是在开始时让每个元素构成一个单元元素的集合，然后按一定顺序将属于同一组的元素所在的集合合并，其间要反复查找一个元素在哪个集合中。其特点是看似并不复杂，但数据量极大，若用正常的数据结构来描述的话，往往在空间上过大，计算机无法承受；即使在空间上勉强通过，运行的时间复杂度也极高，根本就不可能在比赛规定的运行时间（1~3秒）内计算出试题需要的结果，只能用并查集来描述。



什么是并查集？

将编号为 $1 \sim n$ 的 n 个对象划分为不同的集合，在每个集合中，选择其中一个元素进行代表整个集合。

特点：代码短，思路精巧，容易在比赛中出现。

可以高效进行如下操作：

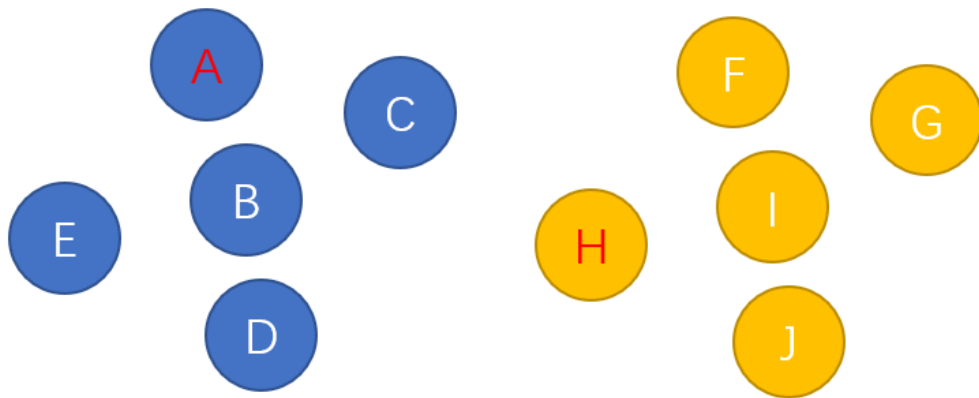
- 查询其中的 a 和 b 是否为同一组。
- 合并元素 a 和元素 b 所在的组。

时间复杂度近乎 $O(1)$ 。



并查集的优势

对于如下图所示的两个集合，如果我们要判断H和A是否在同一个集合中，我们需要遍历A所在的集合，并逐一判断当前节点是否是H节点，直到最后遍历完整个蓝色集合，才能判断出H节点不在这个集合中。





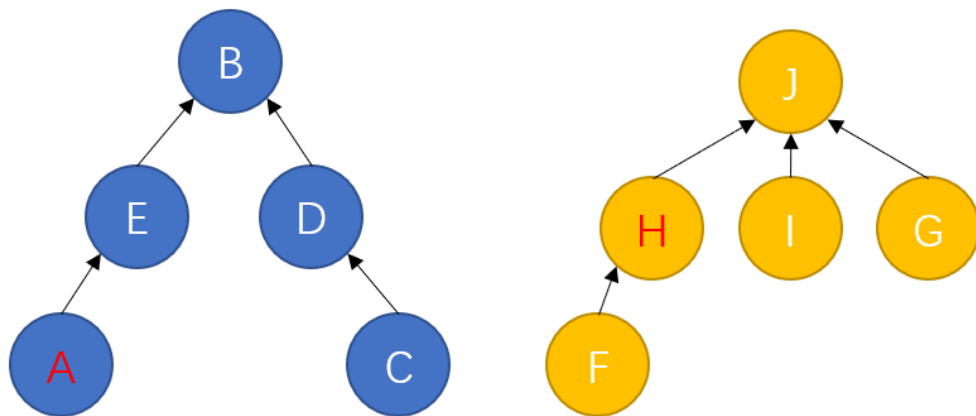
并查集的优势

同样的，如果我们需要合并两个集合，就需要遍历整个黄色的集合，将里面的节点一个一个加入到蓝色集合中。两者都是 $O(n)$ 的复杂度。

但倘若我们在生成集合的时候，就人为地将集合中的元素之间创建某种关联，使它们具有共同的头结点，那么查询和合并的操作将会省时很多。

就拿刚刚的两个集合举例，在创建集合的过程中，为节点之间创建“联系”，形成如下图所示的结构：

可以发现，最终生成的这个结构其实就是一个树形结构。

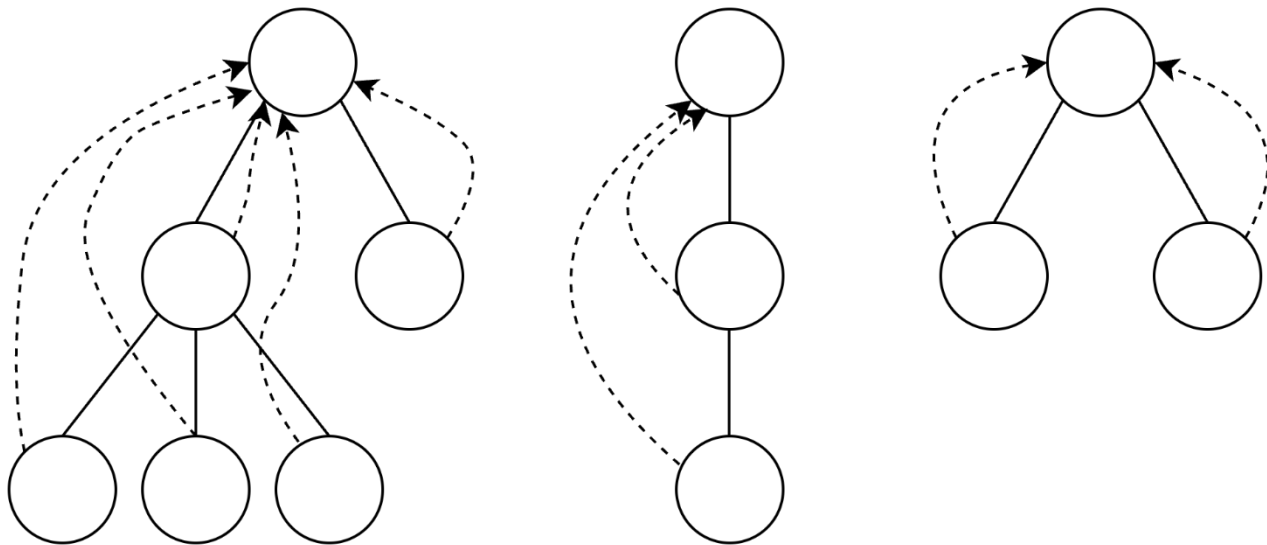


这也就意味着一个集合中的所有节点都可以找到同一个头结点。此时合并和查询操作将变得异常简单。



并查集原理

每个集合用一棵树来表示。树根的编号就是整个集合的编号。每个节点存储他的父节点， $p[x]$ 表示 x 的父节点。

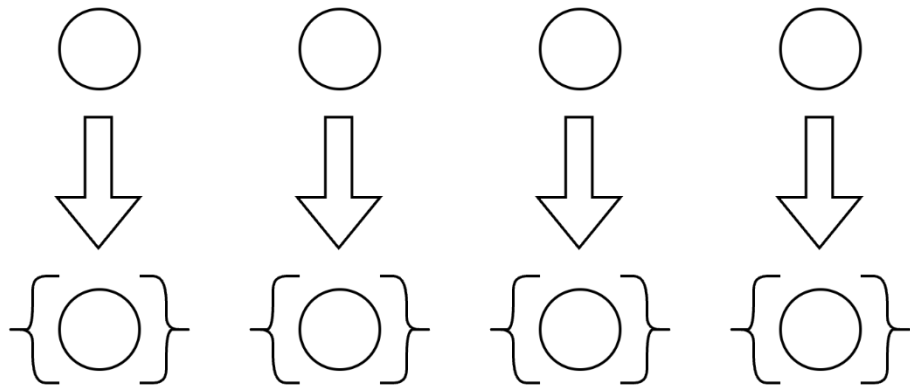




基础操作

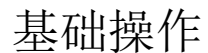
初始化:

对所有 $p[x]=x$ ，自己是自己的根。相当于对每个元素创建一个只有本身的一个集合。

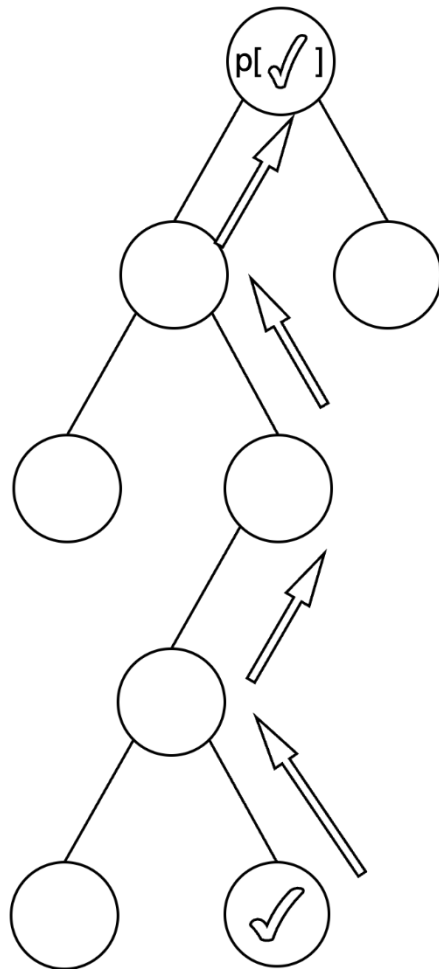


判断是否为根:

则 $p[x]==x$ ，相当于寻找其根节点



求集合的编号：则 $p[x]=x$
就是从当前树一路网上走，走到树根就行了。



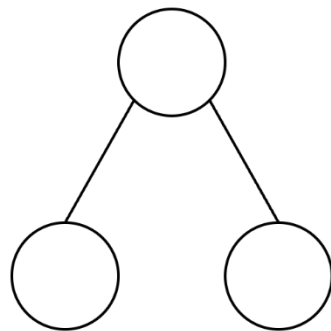
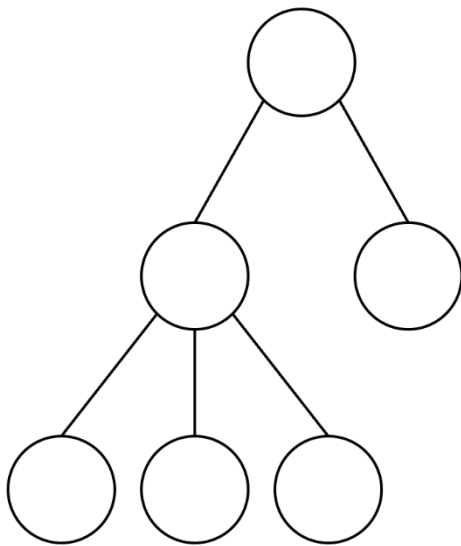


基础操作

合并a和b的集合：
即 $p[a]=b$ 或者 $p[b]=a$

问题在于：如果多次合并，在极端情况下可能会退化成链表。

在层数越少的树中，搜索的越快，
那么如何构建树可以使得树的层数
最小化呢？——按秩合并





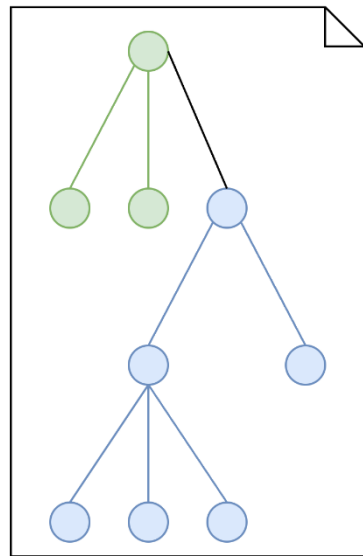
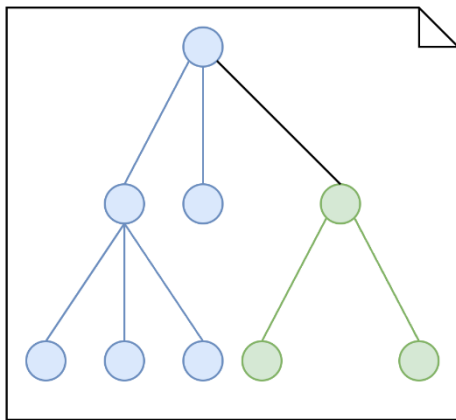
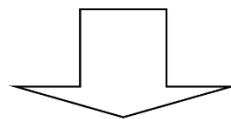
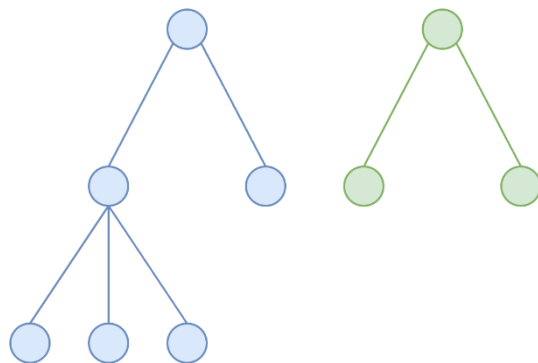
优化方法1-按秩合并

基本思想是使包含较少结点的树的根指向包含较多结点的树的根。

注意：

(1) $p[a] = b$

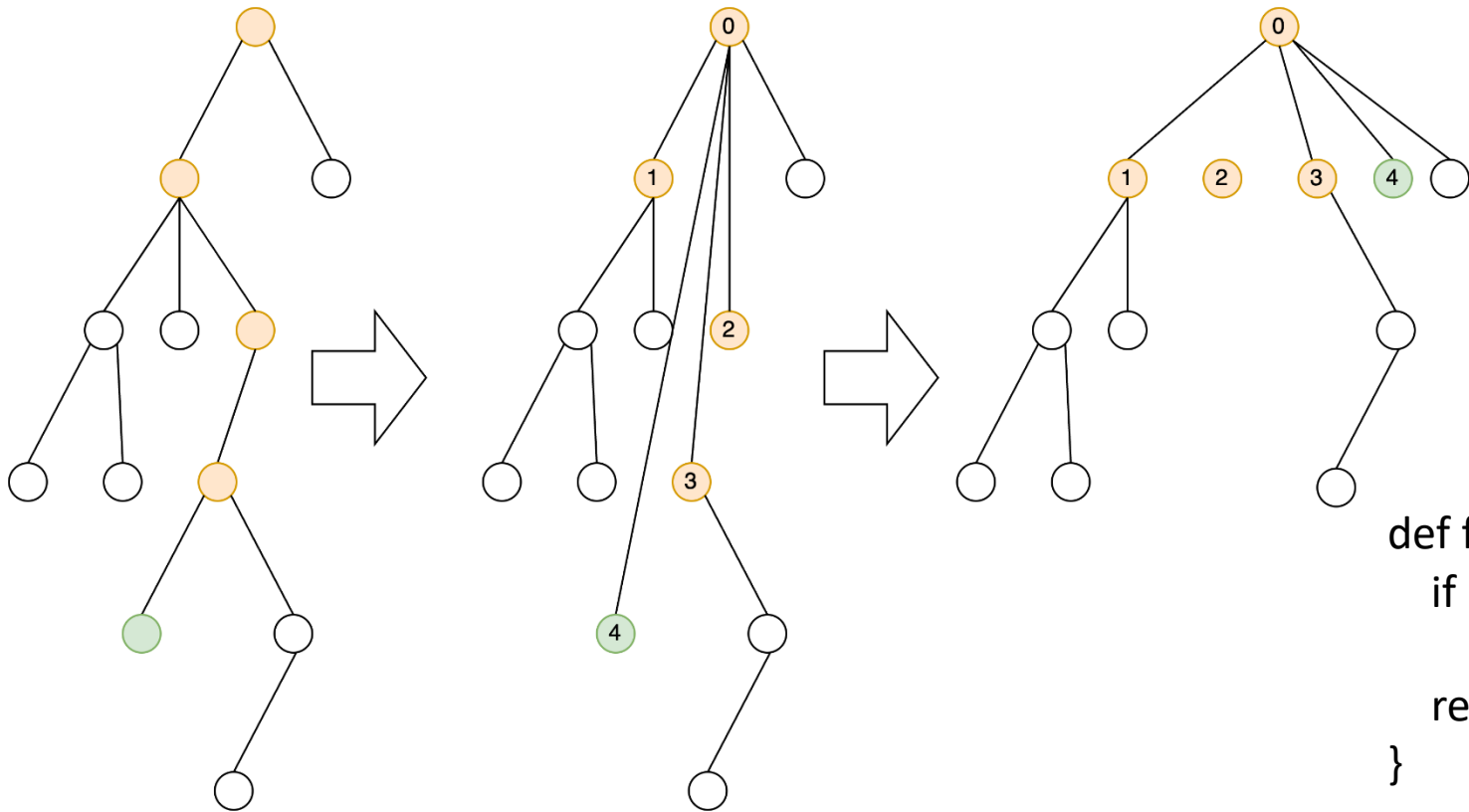
(2) $p[a] = p[b]$





优化方法2-路径压缩

在沿着路径找到根节点后，把路径上的节点直接指向根节点。



```
def find(x) {  
    if 不是根节点:  
        p[x] = find(p[x]);  
    return p[x];  
}
```



例题

动物王国中有三类动物A,B,C，这三类动物的食物链构成了有趣的环形。
A吃B，B吃C，C吃A。

现有N个动物，以 1~N编号。

每个动物都是 A,B,C中的一种，但是我们并不知道它到底是哪一种。

有人用两种说法对这N个动物所构成的食物链关系进行描述：

第一种说法是 1 X Y，表示X和Y是同类。

第二种说法是 2 X Y，表示X吃Y。

此人对N个动物，用上述两种说法，一句接一句地说出K句话，这K句话有的是真的，有的是假的。

当一句话满足下列三条之一时，这句话就是假话，否则就是真话。

- 1.当前的话与前面的某些真的话冲突，就是假话；
- 2.当前的话中X或Y比N大，就是假话；
- 3.当前的话表示X吃X，就是假话。

你的任务是根据给定的N和K句话，输出假话的总数。



例题

输入格式

第一行是两个整数 N 和 K ，以一个空格分隔。

以下 K 行每行是三个正整数 D, X, Y ，两数之间用一个空格隔开，其中 D 表示说法的种类。

若 $D = 1$ ，则表示 X 和 Y 是同类。

若 $D = 2$ ，则表示 X 吃 Y 。

输出格式

只有一个整数，表示假话的数目。

数据范围

$$1 \leq N \leq 50000,$$

$$0 \leq K \leq 100000$$

敬請雅正

