# 第三章、查找

查找即通过给定的条件从数据集合找到所需要的数据

1. 符号表

即以键值对存储的数据结构，查找必须要通过键来查找

* 1. API
     1. 所有对于符号表中的某条具体的数据的操作都必须要通过键。
     2. 符号表是一种抽象数据类型，我们需要指定API，需要实现泛型和迭代
  2. 有序符号表
     1. 作用

对于有序符号表我们有更多可做的操作，例如操作最大最小键、查找键在某一范围的所有数据

* + 1. 实现

所有键都是Comparable的对象，这样可以在插入键值对的时候通过比较来对键排序

* 1. 用例举例
     1. 行为测试用例

符号表可以用于存储键值对，和迭代遍历键值对

* + 1. 性能测试用例

对于符号表处理大型问题，需要能够get()操作更高效，因为无序符号表的get()实现是搜索表中的所有键，对于一个几千条数据的表进行数据处理，其中可能有几百万此搜索。

* 1. 无序链表的顺序查找
     1. get（）的实现

通过比较要查找的键和表中的所有键，如果找到就返回值，否则返回null

* + 1. put（）的实现

也是通过遍历链表，通过equals()来比较要查找的键和表中的所有键，如果匹配成功就更新这个键的值，否则就插入新结点。

* + 1. 顺序查找（基于无序链表）

使用一个私有类来保存结点。get（）会遍历符号表。put（）也会遍历所有键，知道匹配成功就会修改值，匹配失败会添加一个新结点。

* + 1. 实现原理

每个结点保存键和值以及对下一个结点的引用，该对象还会保存一个首结点， 插入新结点的只需覆盖首结点，将原先的首结点的引用加入到新结点中。

* 1. 有序数组中的二分查找
     1. 二分查找（基于有序数组）
        1. 数据结构

将键和值分别用两个数组保存，

* + - 1. get()

使用rank（）找到键保存的位置

* + - 1. put()

如果能找到则更新，否则插入数据，并且将比它大的数据向后移

* + - 1. rank()

使用二分法递归查找键在数组keys的位置，将数组从中间键分为两部分，如果该键大于中间键则继续从右半部分查找，如果该键小于中间键则从左半部分查找，如果相同则直接返回该位置。

* 1. 对二分查找的分析

二分查找的查找速度很快，但是插入很慢。

* 1. 预览

1. 二叉查找树
   1. 基本实现
      1. 数据表示

每个结点含有一个键、值、一条左链接、一条右链接、一个结点计数器

* + 1. 查找

查找从根结点开始，对比查找键和结点键，如果查找键大的话，就递归地从根结点的右子结点开始查找。如果查找键小的话，就递归地从根结点的左子结点开始查找。相等则返回该结点的值

* + 1. 插入

与查找很相似，找到相同的结点就修改值，否则就根据给定键和当前结点键的大小关系来决定继续查找左子树或右子树，直到这个子树是空结点，就返回一个新结点的链接。这就实现了在相应的位置插入一个新结点

* + 1. 递归

对于递归的运行细节我们可以这样理解：对于递归调用前的代码，可以想象沿着树向下走。递归调用后的代码可以想象成沿树往上走

* 1. 分析
  2. 有序性相关的方法和删除操作
     1. 最大键和最小键

取最小键只需查看当前结点是否有左子结点。如果没有则当前结点就是最小键的结点，否则就递归的把左子结点作为当前结点

取最大键则与取最小键同理，只需把左子结点换成右子结点

如果根节点的左链接为空，那么根节点就是最小键。否则递归地从左子树中查找最小键

* + 1. 向下取整和向上取整

如果给定的键key小于根节点的键，那么小于key的最大键一定在根节点的左子树中；如果给定的键大于根结点的键，那么只有当根节点的右子树中存在小于等于key的结点时，目标结点才存在于根节点的右子树中，否则小于key的最大键就是根结点的键。

* + 1. 选择操作

选择操作就是找到排名为k的键（即键中正好右k个小于它的键），如果左子树的结点树t大于k，那么就递归地在左子树中查找排名为k的键。如果t等于k，我们就返回根结点中的键。如果t小于k，我们就递归地在右子树中查找排名为k – t – 1的键

* + 1. 排名

排名就是返回给定键key的排名（即右多少键小于给定键），rank()是select()的逆操作，。如果给定的键与根结点的键相等，则返回左子树的结点总数。如果给定键小于根节点的键，那么则递归地在左子树中查找给定键地排名。如果给定键大于根节点的键，则递归地在右子树中查找排名，并且将给定键在右子树中排名加上根结点左子树的总结点数加一。

* + 1. 删除最大键和最小键

删除最小键我们需要从根结点开始，一直递归地查看左子树，直到根节点没有左子树，这时候我们用根节点的右子树替代根结点（只需要将右子树的引用返回给上一个结点，作为上一个结点的左子树）。

同时递归调用后还需要重新计算结点计数器

* + 1. 删除操作

如果要删除的结点只有一个子结点，我们可以用类似于删除最小键的方法。但是如果有两个子结点就需要用这个结点的后继结点去替代该结点，比如待删除结点有右子结点，那么该结点的后继结点就是右子树的最小结点，并且使用executeDeleteMin()删除右子树的最小结点（后继结点）。然后在递归回调结束之后调整结点计数器就好。如果待删除结点有左子结点，方法也是类似

如果要删除的结点只有一个子结点，我们可以用子结点来代替该结点（只需返回子结点即可），但如果要删除的结点有两个子结点，找到一个后继结点来代替该结点，比如把该结点的右子树的最小结点作为后继结点，因为该结点和该后继结点之间没有任何键，这样不会影响树的结构。只需要一下几步即可完成此替换操作

* + - * 1. 将待删除的结点x保存为t
        2. 将x指向它的后继结点executeMin(t.right)
        3. 将x.right（原本指向空）指向executeDeleteMin(t.right)，也就是删除后所有结点都大于x.key的子树
        4. x.left设为t.left

在递归调用结束后需要重新计算每个结点的结点计数器

* + 1. 范围查找

要实现能够返回给定范围内键的keys()方法，我们首先实现一个遍历二叉查找树的遍历方法—中序遍历。中序遍历需要将所有键按顺序打印出来。要做到这一点我们需要先递归地打印出根结点的左子树所有键，再打印根结点的键，最后递归地打印根结点右子树所有键。

要实现能接受两个参数并且返回给定范围的键的keys()方法，我们需要能够将所有在范围内的键加入到queue中，并且跳过那些不存在给定范围键的子树。

* + 1. 性能分析

1. 平衡查找树

之前所学到的二叉树可能退化成链表

* 1. 2-3查找树

为了保证查找的平衡性，我们需要更加灵活，因此我们允许一个结点保存多个结点。我们将标准二叉树的结点称之为2结点（含有一个键和两个链接），现在我们引入3结点，它含有两个键和三个链接，2结点和3结点的每个连接都对应着由结点的键分割的一个区间。

确切地说，2结点，含有一个键和两个链接，左链接指向小于根结点的键的子树，右链接指向大于根结点键的子树。

3结点，含有两个键和三个链接，左链接指向小于根结点的左键的子树，中链接指向位于根结点的左键和右键之间的子树。右链接指向大于右键的子树。

我们将指向一个空树的链接称之为空链接

一棵完美平衡的2-3查找树中的所有空链接到根结点的距离都是相同的。

* + 1. 查找

将二叉查找树的查找算法一般化可以得到2-3树的查找算法。先将待查找键与根结点的键做对比，和其中任意一个相等则查找命中。否则通过比较的结结果找到相应的链接，再递归地在子树中继续查找

* + 1. 向2-结点插入新键

要在2-3树插入一个新结点，我们可以在树中进行一次未命中的查找，然后将新结点挂在树的底部，但这样无法保持完美平衡。如果未命中的查找结束与一个2结点，我们只需要将这个2结点改为含有2个键的3结点即可。

* + 1. 向一棵只含有一个3结点的树中插入新键

在考虑一般情况前，我们先假设我们需要向一棵只含有一个3-结点的树中插入一个新键。这棵树中有两个键，所以无法再向唯一的结点中插入键了。为了插入新结点，我们先临时的将新键插入到结点中，使之成为一个4-结点，然后我们可以很容易地将一个4-结点转化为含有3个2结点的树，只需要左链接指向左键所创建的2-结点，中键作为根结点键，右链接指向右键所创建的2-结点

* + 1. 向一个父节点为2-结点的3-结点中插入新键

假设一个未命中查找结束于一个3-结点并且父节点是2-结点，这时候我们依然将新键临时保存到3-结点中，使之成为一个4-结点并将其分解。但是我们不会将它转化为3个2-结点，而是将中键放入到父结点中，而父结点中的原中键左边的链接指向原左键所构建的2-结点，父结点中的原中键右边链接指向原右键构建的2-结点。你可以将这次转换看成是将指向原3-结点的链接替换成新父结点的原中键的左右两条链接，并分别指向两个新的2-结点。

* + 1. 向一个父结点为3-结点的3-结点中插入新键

假设一个未命中查找结束于一个3-结点并且父节点是3-结点，这时候我们依然将新键临时保存到3-结点中，使之成为一个4-结点并将其分解。将中键移动到父结点中，但是父结点也是一个3-结点，这样的话，父结点也变成了一个4结点，然后在父结点中进行相同的变换。推广到一般情况，我们就这样递归的一直向上不断分解临时的4-结点，直到遇到一个不需要继续分解的2-结点，或者一个3-结点的根结点。

* + 1. 分解根结点

如果从插入结点到根节点全都是3-结点，我们可以按照向一棵只含有一个3结点的树中插入新键的方式来分解，将一个4-结点分解为3个2-结点。

* + 1. 局部变换

将一个4-结点转换为一棵2-3树可能有6种情况。这个4结点可能是一个根结点，也可能是一个2结点的左子结点或右子结点，也可能是一个3-结点的左子结点或中子结点或右子结点。这些变换都是局部的，除了相关的结点和链接之外不需要再检查树的其他部分，需要指出的是，不光是在树的底部，任何符合这个模式的4-结点都可以使用这种变换。可以将4-结点的中键移动到父结点中，并且这个原中键的左右两个链接分别指向原左键构建成的结点和原右键构建的结点。

* + 1. 全局性质

这些局部变换不会影响全局的有序性和平衡性：任意空链接到根结点的距离都相等。假设变换之前任意空链接到根结点的距离是h，那么变换之后空结点到根结点的距离任然为h。即使是将一个4-结点分解为两个2-结点并且将中键移动到父结点中。因为在变换中树的高度没有发生改变，并且所有空链接任然在最底层。除了将根结点分解为3个2-结点，这个时候树的高度增加了1，所以空链接到根结点的距离变为h+1。

和标准的二叉查找树的向下生长不同，2-3树是向上生长的。它的插入方式只有将一个2-结点变成3-结点，或者一个3-结点临时变成4-结点，再将4-结点分解为2个2-结点以及一个移出的中键，如果这个结点有父结点就将这个移出的中键放到父结点中，如果这个结点是父结点那么移出的中键就单独作为一个根结点。

2-3树在最坏情况下任然有比较好的性能

但是距离真正的实现还有一段差距，因为要处理的情况实在太多，很多操作的实现不方便

* 1. 红黑二叉查找树

红黑树使用由左斜的红链接相连的两个2-结点来表示3-结点。这样的话，红黑树即可以通过转换表示为2-3树，又是一个标准的二叉树。好处在于即可以使用标准二叉的简洁方法来实现各种操作，又可以用2-3树的分解4-结点的方式来进行插入以保持完美的黑链接平衡

* + 1. 替换3-结点

红黑树的思想是用标准的二叉查找树（完全由2-结点构成）和一些额外的信息（替换3-结点）来表示2-3树。我们用由一个红色左链接相连的两个2-结点表示一个3-结点。这中表示法的优点是我们可以用标准二叉树的get（）方法进行查找。对于任意的2-3树，我们只要对结点进行转换，我们都可以立即派生出一棵对应的二叉查找树

* + 1. 一种等价的定义
       1. 任何一个结点不可能与两个红链接相连
       2. 任何一个空链接到根结点之间的黑链接数相同。
       3. 红链接均为左链接
    2. 一一对应

如果将所有的红链接放平的话，我们可以看到所有的空链接到根结点的距离相同。首先，我们可以将红黑树在二叉树和2-3树之间进行转换，如果我们将所有由左斜的红链接相连两个2-结点合并就变成了一个2-3树，反之如果将2-3树的所有3-结点表示为由左斜的红链接相连的两个2-结点，它就变成了一个标准二叉树。红黑树所有空链接到根结点之间的黑链接树相同，因为黑链接就是原本2-3树的普通链接。并且所有结点都不可能同时连接两个红链接。如果我们能够结合二叉树和2-3树的算法，我们就可以同时使用二叉树简洁的查找算法，和2-3树插入的完美平衡。算法

* + 1. 颜色表示
    2. 旋转
    3. 在旋转后重置父结点的链接
  1. 红黑树的插入算法
  2. 删除操作
  3. 红黑树的性质

1. 散列表
2. 1