二分查找BinarySearch及区间查找SearchRange

# 二分查找BinarySearch

## 简单介绍

**二分查找**，又叫**折半查找；是一种效率较高的查找算法。**

### 二分查找的要求：

线性表是有序表，即**表中结点按关键字有序**，并且要用向量作为表的存储结构。不妨设有序表是递增有序的。

**存储结构：**二分查找只适用于顺序存储结构且要求元素按关键字有序排列。

### 适用情况

二分查找只适用**顺序存储结构**。为保持表的有序性，在顺序结构里插入和删除都必须移动大量的结点。因此，**二分查找特别适用于那种一经建立就很少改动、而又经常需要查找的线性表**。

对那些查找少而又经常需要改动的线性表，可采用**链表作存储结构**，进行顺序查找。**链表上无法实现二分查找**。

## 二分查找的基本思想

二分查找的基本思想是：

设R[low…..high]是当前的查找区间

1. 首先确定该区间的**中点位置**：mid = (low+high)/2;
2. 然后将待查的K值与arr[mid]比较：若相等，则查找成功并返回此位置，否则须确定新的查找区间，继续二分查找，具体方法如下：

* 若arr[mid]>K，则由表的有序性可知arr[mid….high]均大于K，因此若表中存在关键字等于K的结点，则该结点必定是在位置mid左边的子表R[low…mid-1]中，故新的查找区间是左子表arr[low…...mid-1]。
* 若arr[mid]<K，则要查找的K必在mid的右子表arr[mid+1……high]中，即新的查找区间是右子表arr[mid+1……high]。下一次查找是针对新的查找区间进行的。

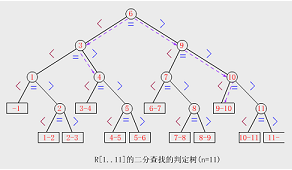
因此，从初始的查找区间R[1..n]开始，每经过一次与当前查找区间的中点位置上的结点关键字的比较，就可确定查找是否成功，不成功则当前的查找区间就缩小一半。这一过程重复直至找到关键字为K的结点，或者直至当前的查找区间为空(即查找失败)时为止。

* 成功的情况下：查找过程走了一条**从判定树的根到被查记录的路径**，比较的次数恰为记录在树中的层数。
* 失败的情况下：查找过程经过了从**判定树的根**到**某个外部节点**的路径，所需比较次数不超过判定树的高度。

二分查找的时间复杂度：**时间复杂度O(log2N)**。

## 二分查找判定树：

二分查找过程可用二叉树来描述：把当前查找区间的中间位置上的数据作为**根**，左子表和右子表分别作为根的左子树和右子树。由此得到的二叉树，称为**描述二分查找的判定树(Decision Tree)或比较树(Comparison Tree)**。



注意：判定树的形态只与表结点个数n相关，而与输入实例中arr[low…..high]的取值无关。

（1）二分查找判定树的组成

①圆结点即树中的内部结点。树中圆结点内的数字表示该结点在有序表中的位置。

②外部结点：圆结点中的所有空指针均用一个虚拟的方形结点来取代，即外部结点。

③树中某结点i与其左(右)孩子连接的左(右)分支上的标记"<"、"("、">"、")"表示：当待查关键字K<R[i].key(K>R[i].key)时，应走左(右)分支到达i的左(右)孩子，将该孩子的关键字进一步和K比较。若相等，则查找过程结束返回，否则继续将K与树中更下一层的结点比较。

（2）二分查找判定树的查找

二分查找就是将给定值K与二分查找判定树的根结点的关键字进行比较。若相等，成功。否则若小于根结点的关键字，到左子树中查找。若大于根结点的关键字，则到右子树中查找。

【例】对于有11个结点的表，若查找的结点是表中第6个结点，则只需进行一次比较；若查找的结点是表中第3或第9个结点，则需进行二次比较；找第1，4，7，10个结点需要比较三次；找到第2，5，8，11个结点需要比较四次。

由此可见，成功的二分查找过程恰好是走了一条**从判定树的根到被查结点的路径**，经历**比较**的关键字次数恰为**该结点在树中的层数**。若查找失败，则其比较过程是经历了一条从判定树根到某个外部结点的路径，所需的关键字比较次数是该路径上内部结点的总数。

【例】待查表的关键字序列为：(05，13，19，21，37，56，64，75，80，88，92)，若要查找K=85的记录，所经过的内部结点为6、9、10，最后到达方形结点"9-10"，其比较次数为3。

实际上方形结点中"i-i+1"的含意为被查找值K是介于R[i].key和R[i+1].key之间的，即R[i].key<K<R[i+1].key。

② 二分查找的平均查找长度

设内部结点的总数为**n=2h-1**，则判定树是深度为**h=lg(n+1)**的满二叉树(深度h不计外部结点)。树中第k层上的结点个数为2k-1，查找它们所需的比较次数是k。因此在等概率假设下，二分查找成功时的平均查找长度为：

ASLbn≈lg(n+1)-1

二分查找在查找失败时所需比较的关键字个数不超过判定树的深度，在最坏情况下查找成功的比较次数也不超过判定树的深度。即为：

 二分查找的最坏性能和平均性能相当接近。

## 二分查找的优缺点

### 二分查找的优点

折半查找的时间复杂度为O(logn)，远远好于顺序查找的O(n)。

### 二分查找的缺点

虽然二分查找的效率高，但是**要将表按关键字排序**。而排序本身是一种很费时的运算。即使采用高效率的排序方法也要**花费O(nlgn)**的时间。

# 二分查找的Java代码

## LeetCode题目：704. Binary Search

难度：Easy

<https://leetcode.com/problems/binary-search/description/>

Given a sorted (in ascending order) integer array nums of n elements and a target value, write a function to search target in nums. If target exists, then return its index, otherwise return -1.

**Example 1**:

Input: nums = [-1,0,3,5,9,12], target = 9

Output: 4

Explanation: 9 exists in nums and its index is 4

**Example 2:**

Input: nums = [-1,0,3,5,9,12], target = 2

Output: -1

Explanation: 2 does not exist in nums so return -1

Note:

You may assume that all elements in nums are unique.

n will be in the range [1, 10000].

The value of each element in nums will be in the range [-9999, 9999].

## Java代码

**递归**实现与**while循环**实现，一般使用**while循环**比较常用。

### 递归实现：

/\*\*

\* 递归方法实现

\*/

public int **binarySearchRecursive**(int[] nums,int low,int high,int target){

if(low > high) return -1;//递归终止条件

int mid = (low + high)/2;

if(target < nums[mid]){

return binarySearchRecursive(nums,**low,mid-1**,target);

}else if(target > nums[mid]){

return binarySearchRecursive(nums,**mid+1,high**,target);

}else{

return mid;

}

}

### **while循环实现BinarySearch(重点掌握)**

public int **binarySearch**(int[] arr,int low,int high,int value){

while(low <= high){

int mid = (low+high)/2;

**if(value < arr[mid]){**

**high = mid - 1;**

**}else if(value > arr[mid]){**

**low = mid + 1;**

**}else{**

**return mid;//返回查找的索引**

**}**

}

return -1;//没有找到

}

# 查找范围BinarySearchRange

## LeetCode题目：

**34. Find First and Last Position of Element in Sorted Array**

**难度：Medium**

Given an array of integers nums sorted in ascending order, find the starting and ending position of a given target value.

Your algorithm's runtime complexity must be in the order of **O(log n)**.

If the target is not found in the array, return **[-1, -1]**.

**Example 1:**

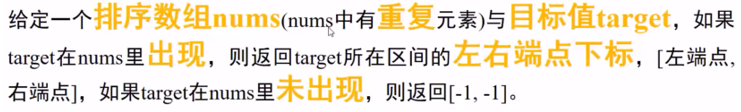
Input: nums = [5,7,7,8,8,10], target = 8

Output: [3,4]

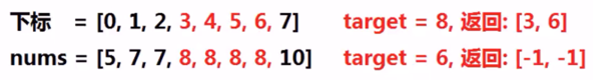
**Example 2:**

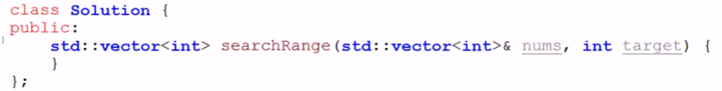
Input: nums = [5,7,7,8,8,10], target = 6

Output: [-1,-1]



示例：





## 思路分析

**大概思路**：要找出target的范围，只需要分别找出左端点和右端点即可。(当然也存在同时寻找左右端点的优秀算法，由于时间关系，这里不深入研究)

在二分查找的基础上修改：当target不等于(大于或小于)nums[mid]时，按照二分查找原则改变搜索范围即可，直到找到target==nums[mid];这里重点对target==nums[mid]情况处理：

第一种思路：判断是否到达端点<端点条件>，若没有，则不断更改搜索范围；

* 查找左端点：**真正左端点有两种情况：①mid==low；②target > nums[mid-1]**。首先判断是否满足这两种情况(或)，若满足，直接返回mid，否则继续向下调整空间，即**high=mid-1**；
* 查找右端点：**真正右端点的两种情况：①mid==high;②target<nums[mid+1]**。首先判断是否满足这两种情况(或)，若满足，直接返回mid，否则继续向上调整空间，即**low=mid+1**；

第二种思路：基于**while循环<非端点条件>**，直到找到左端点或右端点；

* 查找左端点：非左端点情况下，mid依次递减，直到找到左端点，非左端点的条件： while(**--mid >= low &&nums[mid] == target**);跳出while循环代表找到了左端点，return mid+1;即可。
* 查找右端点：非右端点情况下，mid依次递增，直到找到右端点，非右端点的条件：while(++mid <= high && nums[mid] == target);跳出while循环代表找到了右端点，return mid-1;即可。

## Java代码

### 第一种思路：

获取左端点：

public int **binarySearchLeft2**(int[] nums,int low,int high,int target){

while(low <= high){//这里包括等号，与快排不一样，快排不包括等号，二分查找必须包括等号，否则出错

int mid = (low+high)/2;

if(target < nums[mid]){

high = mid - 1;

}else if(target > nums[mid]){

low = mid + 1;

}else{

**if(mid == low||target>nums[mid-1])**

**return mid;**

**high = mid - 1;**

}

}

return -1;//不存在则返回-1

}

查找右端点：

public int **binarySearchRight2**(int[] nums,int low,int high,int target){

while(low <= high){//这里包括等号，与快排不一样，快排不包括等号，二分查找必须包括等号，否则出错

int mid = (low+high)/2;

if(target < nums[mid]){

high = mid - 1;

}else if(target > nums[mid]){

low = mid + 1;

}else{

**if(mid == high||target<nums[mid+1])**

**return mid;**

**low = mid + 1;**

}

}

return -1;//不存在则返回-1

}

### 第二种思路：

查找左端点：

/\*\*

\* 若数组中存在多个相同的值，返回最左边那个值的索引

\*/

public int **binarySearchLeft**(int[] nums,int low,int high,int target){

while(low <= high){

//这里包括等号，与快排不一样，快排不包括等号，二分查找必须包括等号，否则出错

int mid = (low+high)/2;

if(target < nums[mid]){

high = mid - 1;

}else if(target > nums[mid]){

low = mid + 1;

}else{

**while(--mid >= low &&nums[mid] == target);**

**return mid+1;//返回查找的索引**

}

}

return -1;//不存在则返回-1

}

查找右端点：

/\*\*

\*若数组中存在多个相同的值，返回最右边那个值的索引

\*/

public int binarySearchRight(int[] nums,int low,int high,int target){

while(low <= high){

//这里包括等号，与快排不一样，快排不包括等号，二分查找必须包括等号，否则出错

int mid = (low+high)/2;

if(target < nums[mid]){

high = mid - 1;

}else if(target > nums[mid]){

low = mid + 1;

}else{

**while(++mid <= high && nums[mid] == target);**

**return mid-1;//返回查找的索引**

}

}

return -1;//不存在则返回-1

}

### SearchRange

/\*\*

\* 获取范围

\*/

public int[] binarySearchRange(int[] nums,int low,int high,int target){

int[] result = new int[2];

result[0] = binarySearchLeft(nums,low,high,target);

result[1] = binarySearchRight(nums,low,high,target);

return result;

}

### 综合在一起

/\*\*

\* 获取范围2：综合

\*/

public int[] binarySearchRange3(int[] nums,int low,int high,int target){

int low1 = low,high1 = high;

int[] result = {-1,-1};

while(low <= high){

int mid = (low + high)/2;

if(target<nums[mid]){

high = mid -1;

}else if (target > nums[mid]){

low = mid + 1;

}else{

while(--mid>=low && target==nums[mid]);

result[0] = mid + 1;

break;

}

}

low = low1;

high = high1;

while(low <= high){

int mid = (low + high)/2;

if(target < nums[mid]){

high = mid - 1;

}else if(target > nums[mid]){

low = mid + 1;

}else{

while(++mid <= high&&target==nums[mid]);

result[1] = mid-1;

return result;

}

}

return result;

}