**第5章 查找 学习指导**

1、时间

2020年5月05日周二 （1-4班3-4节，5-10班、数学、大数据专业5-6节）

2020年5月07日周四 （1-4班3-4节，5-10班、数学、大数据专业5-6节）

2次课共4学时。

2、内容

电子教材&PPT预习；

腾讯课堂讲解本章重点和难点；

QQ群交流练习题（中间根据需要采用QQ视频或腾讯视频讲解）；

具体安排：

|  |  |  |
| --- | --- | --- |
| 1 | 5月05日 | (1) 查找算法的实现以及各种查找方法的时间性能； |
| 2 | 5月07日 | (2) 在树结构和散列结构上进行查找的基本思想和方法；  (3) 基于关键字查找与基于关键字散列地址查找的本质区别。 |
| 3 | 5月12日 | （备用） |

**3、知识要点**

**3.1 基本概念**

**查找表**是由具有同一类型（属性）的数据元素（记录）组成的集合。

**静态查找表**：仅对查找表进行查找操作，而不改变查找表中的数据元素；

**动态查找表**：对查找表除进行查找操作外，可能还要进行向表中插入数据元素，或删除表中数据元素的表。

查找的同时对表做修改操作（如插入或删除）则相应的表称之为动态查找表，否则称之为静态查找表。

**关键字**：数据元素中某个数据项的值，可以标识一个数据元素。

**主关键字**：可以唯一标识一个数据元素。

**平均查找长度ASL**：衡量查找算法效率优劣的标准是在查找过程中对关键字需要执行的平均比较次数。

B树；

**3.2 查找的分类**

**线性查找的方法**： ·顺序查找：逐个查找，ASL=（n+1）/2；

**二分查找：**取中点int（n/2）比较，若小就比左区间，大就比右区间。用二叉判定树表示。ASL=（∑（每层结点数\*层数））/N.

**分块查找：**要求“分块有序”，将表分成若干块内部不一定有序，并抽取各块中的最大关键字及其位置建立有序索引表。

**二叉排序树（BST）：**二叉排序树是空树或者满足如下性质的二叉树：

1. 若它的左子树非空，则左子树上所有结点的值均小于根结点的值；
2. 若它的右子树非空，则右子树上所有结点的值均大于根结点的值；
3. 左、右子树本身又是一棵二叉排序树。

二叉排序树的插入、建立、删除的算法平均时间性能是O（nlog2n）。

二叉排序树的删除操作可分三种情况进行处理：

1. \*P是叶子，则直接删除\*P，即将\*P的双亲\*parent中指向\*P的指针域置空即可。
2. \*P只有一个孩子\*child，此时只需将\*child和\*p的双亲直接连接就可删去\*p.
3. \*p有两个孩子，则先将\*p结点的中序后继结点的数据到\*p，删除中序后继结点。

**平衡二叉树AVL：**或者是一棵空二叉树，或者具有如下性质的二叉查找树：其左子树和右子树都是高度平衡的二叉树，且左子树和右子树高度之差的绝对值不超过1。

结点的平衡因子 BF (Banlanced Factor) 定义为：结点左子树与右子树的高度之差。AVL中的任意结点的 BF 只可能是 -1，0，+1

**B-树（多路平衡查找树）**：适合在磁盘等直接存取设备上组织动态的查找表，是一种外查找算法。建立的方式是从下向上拱起。

**散列技术：**将结点按其关键字的散列地址存储到散列表的过程称为散列。散列函数的选择有两条标准：简单和均匀。

**哈希表**：散列表（Hash table，也叫哈希表），是根据关键码值(Key value)而直接进行访问的数据结构。也就是说，它通过把关键码值映射到表中一个位置来访问记录，以加快查找的速度。这个映射函数叫做散列函数，存放记录的数组叫做散列表。

**常见的散列函数构的造方法：**

1. 平方取中法：hash=int（（x^2）%100）
2. 除余法：表长为m，hash=x%m
3. 相乘取整法：hash=int（m\*（x\*A-int（x\*A））；A=0.618
4. 随机数法：hash=random（x）。

**处理冲突的方法：**

1. 开放定址法： 一般形式为hi=（h（key）+di）%m1≤i≤m-1，开放定址法要求散列表的装填因子α≤1.
2. 开放定址法类型： ·线性探查法：address=（hash（x）+i）%m；
3. 二次探查法：address=（hash（x）+i^2）%m；
4. 双重散列法：address=（hash（x）+i\*hash（y））%m；
5. 拉链法：是将所有关键字为同义词的结点链接在同一个单链表中。

**拉链法的优点：** 拉链法处理冲突简单，且无堆积现象；链表上的结点空间是动态申请的适于无法确定表长的情况；拉链法中α可以大于1，结点较大时其指针域可忽略，因此节省空间；拉链法构造的散列表删除结点易实现。但当结点规模较小时，用拉链法中的指针域也要占用额外空间，不如开放定址法省空间。

**4、习题**

**一、填空题**

1. 顺序查找法，表中元素可以 存放。
2. 在分块查找方法中，首先查找 ，然后再查找相应的块。
3. 顺序查找、二分查找、分块查找都属于 查找。
4. 查找表所含元素个数在查找阶段是固定不变的。
5. 对于长度为n的线性表，若进行顺序查找，则时间复杂度为 。
6. 对于长度为n的线性表，若采用二分查找，则时间复杂度为： 。
7. 理想情况下，在散列表中查找一个元素的时间复杂度为： 。
8. 在关键字序列（7，10，12，18，28，36，45，92）中，用二分查找法查找关键字92，要比较 次才找到。
9. 设有100个元素，用二分查找时，最大的比较次数是 次。
10. 对二叉排序树进行查找的方法是用待查的值与根结点的键值进行比较，若比根结点小，则继续在 子树中查找。
11. 二叉排序树是一种 查找表。
12. 哈希表是按 存储方式构造的存储结构
13. 哈希法既是一种存储方法，又是一种 方法。
14. 散列表的查找效率主要取决于散列表造表时选取的散列函数和处理 的方法。
15. 设散列函数H和键值k1，k2，若k1≠k2，且H（k1）=H（k2），则称这种现象为 。
16. 处理冲突的两类主要方法是开放定址法和 。
17. 查找法的平均查找长度与元素个数n无关。
18. 在哈希函数H(key)=key % P中，P一般应取 。
19. 在查找过程中有插入元素或删除元素操作的，称为 查找。
20. 各结点左右子树深度之差的绝对值至多为 的二叉树称谓平衡二叉树。
21. 在10阶B—树中根结点所包含的关键码个数最多为 ，最少为 。
22. 一棵5阶B—树中，除根结点外，每个结点的子树树目最少为 ，最多为 。
23. 对于包含n个关键码的m阶B—树，其最小高度是 ，最大高度是 。
24. 在一棵B—树中删除关键码，若最终引起树根结点的合并，则新树比原树的高度减少 层。
25. 在一棵高度为h的B—树中，叶子结点处于第 层，当向该B—树中插入一个新关键码时，为查找插入位置需读取 个结点。
26. 对于长度为n的线性表，若采用分块查找（假定总块数和每块长度均接近 ，用顺序查找确定所在块），则时间复杂性为 。

**二、选择题**

（1）查找表是以（ ）为查找结构。

A．集合 B．图 C．树 D．文件

（2）顺序查找法适合于存储结构为（ ）的线性表。

A．散列存储 B．顺序存储或链接存储

C．压缩存储 D．索引存储

（3）在表长为ｎ的链表中进行线性查找，它的平均查找长度为（ ）。

Ａ. ASL＝ｎ; Ｂ. ASL＝(n+1)/２;

Ｃ. ASL＝＋１; Ｄ. ASL≈log2ｎ

（4）对线性表进行二分查找时，要求线性表必须 （ ）。

A．以顺序方式存储 B．以链接方式存储，且结点按关键字有序排序

C．以链接方式存储 D．以顺序方式存储，且结点按关键字有序排序

（5）衡量查找算法效率的主要标准是（ ）。

A．元素个数 B. 平均查找长度 C．所需的存储量 D．算法难易程度

（6）如果要求一个线性表既能较快地查找，又能适应动态变化的要求，可以采用（ A ）查找方法。

A．分块 B．顺序 C．二分 D．散列

（7） 链表适用于（ ）查找。

A．顺序 B．二分 C．随机 D．顺序或二分

（8）一个有序表为{1，3，9，12，32，41，45，62，75，77，82，95，100}，当二分查找值为82的结点时，( )次比较后查找成功。

A．2 B．3 C．4 D．5

（9）二分查找有序表{4，6，10，12，20，30，50，70，88，100},若查找表中元素58，则它将依次与表中（ ）比较大小，查找结果是失败。

A．30，88，70，50 B．20，70，30，50 C．20，50 D．30，88，50

（10）对有14个元素的有序表A[1**..**14]作二分查找，查找元素A[4]时的被比较元素依次为（ ）。

A．A[1]，A[2]，A[3]，A[4] B．A[1]，A[14]，A[7]，A[4]

C．A[7]，A[3]，A[5]，A[4] D．A[7]，A[5]，A[3]，A[4]

（11）有一个长度为12的有序表，按二分查找法对其进行查找，在表内各元素等概率情况下查找成功所需的平均比较次数为（ ）。

A．35／12 B．37／12 C．39／12 D．43／12

（12）采用分块查找时，若线性表共有625个元素，查找每个元素的概率相等，假设采用顺序查找来确定结点所在的块时，每块分（ ）个结点最佳。

A．6 B．10 C．25 D．625

（13）下列（ ）不是利用查找表中数据元素的关系进行查找的方法。

A．平衡二叉树 B．有序表的查找

C. 散列查找 D．二叉排序树的查找

（14）设哈希表长m=14，哈希函数H（key）=key％11。表中已有4个结点：

addr(15)=4

addr(38)=5

addr(61)=6

addr(84)=7

其余地址为空。如用二次探测再散列处理冲突，关键字为49的结点的地址是（ ）。

A．8 B．3 C．5 D．9

（15）对包含n个元素的散列表进行查找，平均查找长度为（ ）。

A．O(n2) B．O(log2n) C. O(n) D．不直接依赖于n

（16）冲突指的是（ ）。

A．两个元素具有相同序号 B. 两个元素的键值不同

C．不同键值对应相同的存储地址 D．两个元素的键值相同

（17）在查找过程中，不做增加、删除或修改的查找称为（ ）。

A．静态查找 B. 内创造 C．动态查找 D．外查找

（18）已知8个元素为{34，76，45，18，26，54，92，65}，按照依次插入结点的方法生成一棵二叉树，最后两层上结点的总数为（ ）。

A．1 B．2 C. 3 D．4

（19）不可能生成下图二叉排序树的关键字的序列是（ ）。

A． 4 5 3 1 2 B．4 2 5 3 1 C．4 5 2 1 3 D．4 2 3 1 5

（20）动态查找包括（ ）查找。

A．顺序表 B. 二叉排序树

C．有序表 D．索引顺序表

（21）当向一棵m阶的B-树做插入操作时，若一个结点中的关键字个数等于（ ），则必须分裂为两个结点。

A.m B. m-1 C.m+1 D. m/2

（22）在一个5阶的B-树上，每个非终端结点所含的子树数最少为（ ）。

A. 2 B.3 C.4 D.5

**三、应用题**

（1） 对于给定结点的关键字集合K={5，7，3，1，9，6，4，8，2，10}，

1）试构造一棵二叉排序树；

2）求等概率情况下的平均查找长度ASL。

（2）对于给定结点的关键字集合K={10，18，3，5，19，2，4，9，7，15}，

1）试构造一棵二叉排序树；

2）求等概率情况下的平均查找长度ASL。

（3）将数据序列：25，73，62，191，325，138，依次插入下图所示的二叉排序树，并画出最后结果。

（4）对于给定结点的关键字集合K={1，12，5，8，3，10，7，13，9}，

1）试构造一棵二叉排序树；

2）在二叉树排序BT中删除“12”后的树结构：

（5）对于给定结点的关键字集合K={34，76，45，18，26，54，92，38}，

1）试构造一棵二叉排序树；

2）求等概率情况下的平均查找长度ASL。

（6）对于给定结点的关键字集合K={4，8，2，9，1，3，6，7，5}，

1）试构造一棵二叉排序树；

2）求等概率情况下的平均查找长度ASL。

解：

（7）画出对长度为10的有序表进行折半查找的判定树，并求其等概率时查找成功的平均查找长度。

（8）二叉排序树如图所示，分别画出：

1）画出删除关键字15以后的二叉树，并要求其平均查找长度尽可能小；

2）在原二叉排序树（即没有删除15）上，插入关键字20。

（9）给定结点的关键字序列为：19，14，23，1，68，20，84，27，55，11，10，79。

设散列表的长度为13，散列函数为：H（K）=K % 13。试画出线性探测再散列解决冲突时所构造的散列表，并求出其平均查找长度。

（10）给定结点的关键字序列为：47，7，29，11，16，92，22，8，3，哈希表的长度为11。

设散列函数为：H（K）=K % 11。试画出平方探测再散列解决冲突时所构造的散列表，并求出其平均查找长度。

（11）给定结点的关键字序列为：19，14，23，1，68，20，84，27，55，11，10，79。

设散列表的长度为13，散列函数为：H（K）=K % 13。试画出链地址法解决冲突时所构造的哈希表，并求出其平均查找长度。

（12）给定结点的关键字序列为：47，7，29，11，16，92，22，8，3，哈希表的长度为11。

设散列函数为：H（K）=K %11。试画出链地址法解决冲突时所构造的哈希表，并求出其平均查找长度。

**四、算法设计题**

（1）设单链表的结点是按关键字从小到大排列的，试写出对此链表进行查找的算法。如果查找成功，则返回指向关键字为x的结点的指针，否则返回NULL。

（2）试设计一个在用开放地址法解决从突的散列表上删除一个指定结点的算法。

（3）设给定的散列表存储空间为H[1-m]，每个单元可存放一个记录，H[i]的初始值为零，选取散列函数为H(R.key)，其中key为记录R的关键字，解决冲突的方法为线性控测法，编写一个函数将某记录R填入到散列表H中。