# 查找算法平均查找长度的计算方法

秦玉平1,王丽君2,刘 伟3

(1. 渤海大学 工学院, 辽宁 锦州 121013; 2. 渤海大学 大学计算机教学部 辽宁 锦州 121013; 3. 渤海大学 旅游学院, 辽宁 锦州 121013)

摘 要:给出了常用查找算法平均查找长度的计算方法,包括查找成功和查找失败平均查找 长度的计算,并通过实例进行了解析,便于学习者学习和掌握查找算法,同时为应用者选择查找 算法提供依据。

关键词:查找;平均查找长度;等概率

中图分类号:TP 391.41 文献标识码:A 文章编号:1673 - 0569(2011)04 - 0353 - 05

### 0 引言

查找就是在含有若干记录的表中找出关键字值与给定值相同的记录。若表中存在这样的记录,则查找成功,返回所找到记录的信息或记录在表中的位置;否则查找失败,返回空记录或空指针。查找表有静态查找表、动态查找表和散列表三种<sup>[1]</sup>。每种查找表都有相应的查找算法。一般的查找过程都是对查找表内的记录按一定顺序进行关键字的比较,所以一个查找算法效率的优劣主要取决于查找过程中的比较次数。绝对比较次数计算起来比较繁琐,所以一般用平均比较次数(也称为平均查找长度)作为衡量一个查找算法效率的标准。平均查找长度 ASL(Average search length)定义为<sup>[2]</sup>

$$ASL = \sum_{i=1}^{n} p_i c_i \tag{1}$$

其中,n 为查找表中记录的个数, $p_i$  是在查找表中查找第i 个记录的概率。为了简单起见,通常认为查找每个记录的概率是相等的,即, $p_i = 1/n(1 \le i \le n)$ ; $c_i$  是查找第i 个记录所需的比较次数。计算给定查找算法查找成功或查找失败时的平均查找长度是各类数据结构课程考试常考的题目<sup>[3-5]</sup>,也是实际中选择查找方法的依据<sup>[6-8]</sup>,但所有教材和指导书介绍的都不够全面。为此,本文给出了常用查找算法查找成功和查找失败时平均查找长度的计算方法,并通过实例进行解析。

### 1 常用查找算法平均查找长度的计算方法

#### 1.1 顺序查找法平均查找长度

顺序查找法要求查找表顺序存储或链式存储。顺序查找的过程为:从表的一端开始,依次将表中元素 关键字与给定的值比较,如果表中某个记录的关键字值与给定值相等,则查找成功;若到表的另一端后,仍 然没有找到关键字与给定值相等的记录,则查找失败。

若查找表有n个记录,且采用顺序查找法查找一个记录,则在等概率情况下查找成功时的平均查找长度为

收稿日期:2011-09-30.

基金项目:辽宁省教育厅高等教育教学改革研究项目(2009B145).

作者简介:秦玉平(1965-),男,博士,教授,硕士生导师,从事计算机教学和科研工作.

$$ASL_{succ} = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{n} \times \frac{n(n+1)}{2} = \frac{n+1}{2}$$
 (2)

若设监视哨,则在等概率情况下查找失败时的平均查找长度为

$$ASL_{fail} = n + 1 \tag{3}$$

若不设监视哨,则在等概率情况下查找失败时的平均查找长度为

$$ASL_{fail} = n \tag{4}$$

若查找表为有序表,则在等概率情况下查找成功时的平均查找长度与无序表相同,查找失败时的平均 查找长度为

$$ASL_{fail} = \frac{1}{n+1} \left( \sum_{i=1}^{n} i + n \right) = \frac{n}{2} + \frac{n}{n+1}$$
 (5)

#### 1.2 二分查找法平均查找长度

二分查找法要求查找表按顺序存储,且表中记录按关键字有序。二分查找的过程为:首先确定查找表中点位置,然后用待查值与查找表中间位置上记录的关键字值比较,若相等,则查找成功;否则,若待查值小,则在查找表的前半部分继续进行二分法查找,否则在查找表的后半部分继续进行二分法查找。重复上述的过程,直到查找成功或失败。

若查找表有n个记录,且采用二分查找法查找一个记录,则在等概率情况下查找成功时的平均查找长度为

$$ASL_{succ} = \frac{1}{n} \sum_{i=1}^{n} c_i \tag{6}$$

其中 $,c_i$  为第i条记录在二分查找判定树中的层数。

在等概率情况下查找失败时的平均查找长度为

$$ASL_{fail} = \frac{1}{n+1} \left( \sum_{i=1}^{n} \alpha_i l_i + \sum_{i=1}^{n} \beta_i r_i \right)$$
 (7)

其中, $l_i$  是第 i 条记录在二分查找判定树中的层数,若该结点的左孩子为空,则  $\alpha_i = 1$ ,否则  $\alpha_i = 0$ ; $r_i$ 是第 i 条记录在二分查找判定树中的层数,若该结点的右孩子为空,则  $\beta_i = 1$ ,否则  $\beta_i = 0$ 。

例如,对给定的11个数据元素的有序表(2,3,10,15,20,25,28,29,30,35,40),采用二分查找,其二分查找判定树如图1所示。

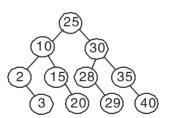


图 1 二分查找判定树

等概率情况下查找成功的平均查找长度为

$$ASL_{succ} = \frac{1 \times 1 + 2 \times 2 + 3 \times 4 + 4 \times 4}{11} = \frac{33}{11} = 3$$

等概率情况下查找失败的平均查找长度为

$$ASL_{fail} = \frac{3 \times 4 + 4 \times 8}{12} = \frac{44}{12} = \frac{11}{3}$$

#### 1.3 在二叉排序树上查找的平均查找长度

二叉排序树的中序遍历序列是一个按关键字递增的有序序列。在二叉排序树上进行查找,是一个从根结点开始,沿某一个分支逐层向下进行比较判断是否相等的过程。即,当二叉排序树不空时,首先将给定值与根结点关键字值比较,若相等,则查找成功;否则,根据给定值与根结点关键字之间的大小关系,在

左子树或右子树上继续进行查找,直到找到,或左子树或右子树为空。由查找过程可知,在二叉排序树上进行查找,若查找成功,则是走了一条从根结点到待查记录的路径;若查找失败,则是走了一条从根结点到某个叶子的路径。

若查找表有n个记录,在二叉排序树上查找一个记录,则在等概率情况下查找成功时的平均查找长度为

$$ASL_{succ} = \frac{1}{n} = \sum_{i=1}^{n} c_i \tag{8}$$

其中, $c_i$  为第 i 条记录在二叉排序树中的层数。

在等概率情况下查找失败时的平均查找长度为

$$ASL_{fail} = \frac{1}{n+1} \left( \sum_{i=1}^{n} \alpha_{i} l_{i} + \sum_{i=1}^{n} \beta_{i} r_{i} \right)$$
 (9)

其中, $l_i$  是第 i 条记录在二叉排序树中的层数,若该结点的左孩子为空,则  $\alpha_i = 1$ ,否则  $\alpha_i = 0$ ; $r_i$ 是第 i 条记录在二叉排序树中的层数,若该结点的右孩子为空,则  $\beta_i = 1$ ,否则  $\beta_i = 0$ 。

例如,输入一个正整数序列(53,17,12,66,58,70,87,25,56,60),按次序构造的二叉排序树如图 2 所示。

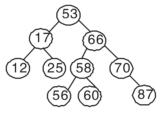


图2 二叉排序树

在等概率情况下查找成功的平均查找长度为

$$ASL_{succ} = \frac{1 \times 1 + 2 \times 2 + 3 \times 4 + 4 \times 3}{10} = \frac{29}{10}$$

在等概率情况下查找失败的平均查找长度为

$$ASL_{fail} = \frac{3 \times 5 + 4 \times 6}{10} = \frac{39}{11}$$

### 1.4 散列查找法平均查找长度

散列表的查找过程和造表过程一致。查找过程为:对于待查值,计算散列地址,若散列地址内为空标记,则查找失败;若待查值与该地址内记录关键字相等,则查找成功。否则,求下一地址,直至散列地址为空标记,或已搜索了表中的所有单元(查找不成功),或待查值与该地址内记录关键字相等(查找成功)为止。决定散列表平均查找长度的因素主要有散列函数、处理冲突的方法和查找表记录数。

若查找表有n个记录,散列函数为H(k) = k%p,采用线性探测再散列法处理冲突,则等概率条件下查找成功时的平均查找长度为

$$ASL_{succ} = \frac{1}{n} \sum_{i=1}^{n} c_i \tag{10}$$

其中 $.c_i$  为查找第i 条记录的比较次数。

等概率条件下查找失败时的平均查找长度为

$$ASL_{fail} = \frac{1}{p} \sum_{i=1}^{p} c_i \tag{11}$$

其中 $,c_i$  为从散列地址i 开始线性探测,直到散列地址为空标记时的探测次数。

例如,若哈希表的地址范围为 $0 \sim 17$ ,散列函数为:H(K) = K% 16,用线性探测再散列法处理冲突,对关键字序列(10,24,32,17,31,30,46,47,40,63,49),散列表如表1 所示。

· · · · · · · · · · · · · · · · · · ·																		
散列地址	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
关键字	32	17	63	49					24	40	10				30	31	46	47
成功比较次数	1	1	6	3					1	2	1				1	1	3	3
失败比较次数	5	4	3	2	1	1	1	1	4	3	2	1	1	1	9	8		

等概率情况下查找成功时的平均查找长度为

$$ASL_{succ} = \frac{1+1+6+3+1+2+1+1+1+3+3}{11} = \frac{23}{11}$$

等概率情况下查找失败时的平均查找长度为

$$ASL_{fail} = \frac{5+4+3+2+1+1+1+1+4+3+2+1+1+1+9+8}{16} = \frac{47}{16}$$

若查找表有n个记录,散列函数为H(k) = k%p,采用链地址法处理冲突,则查找成功时的平均查找长 度为

$$ASL_{succ} = \frac{1}{n} \sum_{i=1}^{n} c_i \tag{12}$$

其中 $,c_i$  为查找第i 条记录的比较次数。

对用拉链法求查找失败时的平均查找长度有两种观点:一种认为比较到空指针算失败,等概率条件下 查找失败时的平均查找长度为

$$ASL_{fail} = \frac{p+n}{p} \tag{13}$$

另一种认为只有和关键字比较才计算比较次数,而和空指针比较不计算,等概率条件下查找失败时的 平均查找长度为

$$ASL_{fail} = \frac{n}{p} \tag{14}$$

例如,若哈希函数 H(k) = 3K%11,散列地址空间为  $0 \sim 10$ ,对关键字序列(32,13,49,24,38,21,4, 12)按链地址法解决冲突。散列表示意图如图 3 所示。

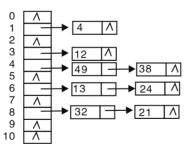


图 3 散列表示意图

查找 4、12、49、13 和 32 时均比较一次, 查找 38、24 和 21 时均比较 2 次, 等概率情况下查找成功时的 平均查找长度为

$$ASL_{succ} = \frac{5 \times 1 + 3 \times 2}{8} = \frac{11}{8}$$

若认为比较到空指针算失败,则哈希地址0、2、5、7、9和10均比较1次失败,哈希地址1和3均比较2 次失败,其余哈希地址均为比较3次失败,因此,等概率情况下查找失败时的平均查找长度

$$ASL_{fail} = \frac{6 \times 1 + 2 \times 2 + 3 \times 3}{11} = \frac{19}{11}$$

若认为只有和关键字比较才算比较次数,而和空指针比较不计算,则等概率情况下查找失败时的平均 查找长度为

$$ASL_{fail} = \frac{8}{11}$$

### 2 结束语

平均查找长度是判断查找方法的一个重要指标,本文给出了顺序查找法、二分查找法、二叉排序树查 找法和散列查找法的平均查找长度的计算方法,每种算法包括查找成功和查找失败平均查找长度的计算, 并通过实例进行了解析,为学习者学习和掌握查找算法提供了帮助,同时为应用者选择查找算法提供了依据。

### 参考文献:

- [1]秦玉平,马靖善.《数据结构(C语言版)》[M].北京:清华大学出版社,2005.
- [2]严蔚敏,吴伟民.《数据结构》(C语言版)》[M].北京:清华大学出版社,1997.
- [3]李春葆.《数据结构考研指导》[M].北京:清华大学出版社,2003.
- [4]李春葆等.《数据结构程序设计题典》[M].北京:清华大学出版社,2002.
- [5]李春葆等.《数据结构习题与解析(C语言篇)》[M].北京:清华大学出版社,2002.
- [6]徐士良.《实用数据结构》[M]. 北京:清华大学出版社,2000.
- [7]徐孝凯.《数据结构辅导与提高实用教程》[M].北京:清华大学出版社,2003.
- [8]严蔚敏,吴伟民.《数据结构题集(c语言版)》[M].北京:清华大学出版社,2008.

## Calculations of search algorithms for average search length

QIN Yu - ping<sup>1</sup>, WANG Li - jun<sup>2</sup>, LIU Wei<sup>3</sup>

- (1. College of Engineering, Bohai University, Jinzhou 121013, China;
- 2. Institute of College Computer Science, Bohai University, Jinzhou 121013, China;
  - 3. Tourism college, Bohai University, Jinzhou 121013, China)

**Abstract:** The commonly – used calculations are provided for search algorithms for average search length, including the successful calculations and the unsuccessful ones. The calculations are illustrated by detailed examples for the reference of learners and users.

**Key words**: search; average search length; equiprobability