



数据结构与算法（十）

张铭 主讲

采用教材：张铭，王腾蛟，赵海燕 编写
高等教育出版社，2008.6（“十一五”国家级规划教材）

<http://www.jpku.pku.edu.cn/pkujpk/course/sjjg>



第十章 检索

- 10.1 线性表的检索
- 10.2 集合的检索
- 10.3 散列表的检索
- 总结



基本概念

- 检索

在一组记录集合中找到关键码值等于给定值的某个记录，或者找到关键码值符合特定条件的某些记录的过程

- 检索的**效率**非常重要

- 尤其对于大数据量
- 需要对数据进行**特殊的存储处理**



提高检索效率的方法

- 预排序

■ 排序算法本身比较费时
 ■ 把数据组织到三个表中
 ■ 根据关键字的值确定表中记录的位置

- 建立索引

■ 缺点：
 ■ 检索时充分利用辅助索引信息
 ■ 牺牲一定的空间出现重复关键字
 ■ 从而提高检索效率

- 散列技术

- 当散列方法不适合于基于磁盘的应用程序时，我们可以选择 B 树方法



平均检索长度 (ASL)

- 关键码的比较：检索运算的主要操作
- 平均检索长度(Average Search Length)
 - 检索过程中对关键码的平均比较次数
 - 衡量检索算法优劣的时间标准

$$ASL = \sum_{i=1}^n P_i C_i$$

■ P_i 为检索第 i 个元素的概率

■ C_i 为找到第 i 个元素所需的关键码值与给定值的比较次数



检索算法评估的其他问题

- 衡量一个检索算法还需要考虑
 - 算法所需的存储量
 - 算法的繁杂性
 - ...



思考

- 假设线性表为 (a, b, c) 检索 a 、 b 、 c 的概率分别为 0.4 、 0.1 、 0.5
 - 顺序检索算法的平均检索长度是多少？（即平均需要多少此次比较给定值与表中关键码值才能找到待查元素）



数据结构与算法

谢谢聆听

国家精品课“数据结构与算法”

<http://www.jpk.pku.edu.cn/pkujpk/course/sjjg/>

张铭, 王腾蛟, 赵海燕

高等教育出版社, 2008. 6. “十一五”国家级规划教材