

7.4散列表（上）

散列表的基本概念

- 散列函数:一个把查找表中的关键字映射成该关键字对应的地址的函数,记为 $\text{Hash}(\text{key}) = \text{Addr}$
- 冲突：散列函数可能会把两个或两个以上的不同关键字映射到同一地址
- 散列表:根据关键字而直接进行访问的数据结构
- 对散列表进行查找的时间复杂度为 $O(1)$

散列函数的构造方法

注意

- 散列函数的定义域必须包含全部需要存储的关键字，而值域的范围则依赖于散列表的大小或地址范围
- 散列函数计算出来的地址应该能等概率、均匀地分布在整个地址空间中，从而减少冲突的发生
- 散列函数应尽量简单，能够在较短的时间内计算出任一关键字对应的散列地址

散列函数

直接定址法

- 直接取关键字的某个线性函数值为散列地址
- 散列函数为 $H(\text{key}) = \text{key}$ 或 $H(\text{key}) = a * \text{key} + b$ a 和 b 是常数
- 特点
 - 计算最简单,且不会产生冲突
 - 适合关键字的分布基本连续的情况
 - 若关键字分布不连续,空位较多,则会造成存储空间的浪费

除留余数法

- 假定散列表表长为 m ,取一个不大于 m 但最接近或等于 m 的质数 p ,利用以下公式把关键字转换成散列地址
- 散列函数为 $H(\text{key}) = \text{key} \% p$
- 特点
 - 最简单、最常用的方法
 - 关键是选好 p ,使得每个关键字通过该函数转换后等概率地映射到散列空间上的任一地址,从而尽可能减少冲突的可能性

数字分析法

- 设关键字是 r 进制数,而 r 个数码在各位上出现的频率不一定相同,可能在某些位上分布均匀一些,每种数码出现的机会均等
- 而在某些位上分布不均匀,只有某几种数码经常出现,此时应选取数码分布较为均匀的若干位作为散列地址
- 特点
 - 适合于已知的关键字集合,若更换了关键字
 - 则需要重新构造新的散列函数

平方取中法

- 取关键字的平方值的中间几位作为散列地址
- 特点
 - 散列地址分布比较均匀
 - 适用于关键字的每位取值都不够均匀或均小于散列地址所需的位数