



哈希查找 | 基本概念

学号范围: XX000 ~ XX999



查找学号：17138?



- 顺序查找: $O(n)$, 平均约比较500次
- 二分查找: $O(\log n)$, 平均约比较10次

有没有效率更高的算法?

取给定学号的后三位，**不需要经过比较**，便可**直接**从查找表中**找到**给定学生的记录。



哈希函数定义

一般情况下，需在关键字与记录在表中的存储位置之间建立一个函数关系，以 $H(\text{key})$ 作为关键字为 key 的记录在表中的位置，通常称这个函数 $h(\text{key})$ 为**哈希函数**。



1) 哈希函数是一个**映象**，即：**将关键字的集合映射到某个地址集合上**，它的设置很灵活，只要这个地址集合的大小不超出允许范围即可；

- 1) 哈希函数是一个**映象**，即：**将关键字的集合映射到某个地址集合上**，它的设置很灵活，只要这个地址集合的大小不超出允许范围即可；
- 2) 由于哈希函数是一个**压缩映象**，因此，在一般情况下，很容易产生“冲突”现象，即： **$\text{key1} \neq \text{key2}$ ，而 $h(\text{key1}) = h(\text{key2})$** 。

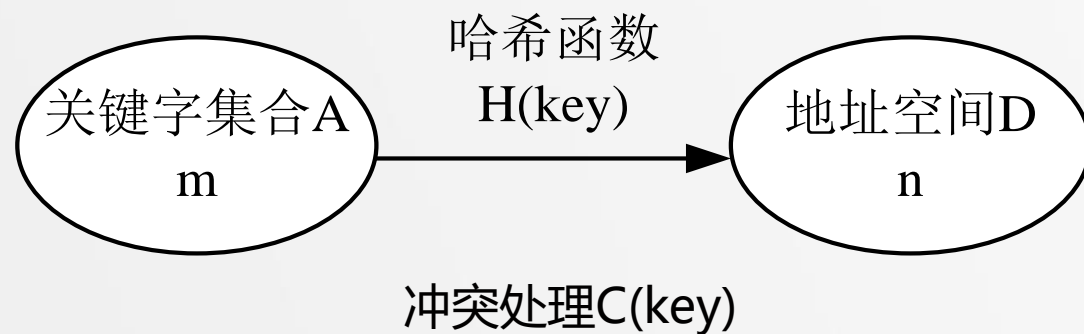
$$H(\text{key}) = \text{key} \% 10$$

- 1) 哈希函数是一个**映象**，即：**将关键字的集合映射到某个地址集合上**，它的设置很灵活，只要这个地址集合的大小不超出允许范围即可；
- 2) 由于哈希函数是一个**压缩映象**，因此，在一般情况下，很容易产生“冲突”现象，即： **$\text{key1} \neq \text{key2}$ ，而 $h(\text{key1}) = h(\text{key2})$** 。
- 3) **很难**找到一个不产生冲突的哈希函数。一般情况下，**只能**选择恰当的哈希函数，使冲突尽可能少地产生。

因此，哈希查找需要做两方面事情：选择一个“**好**”的哈希函数；提供一种“**处理冲突**”的方法。

哈希表

根据设定的**哈希函数 $H(\text{key})$** 和提供的**处理冲突的方法**，将一组关键字**映射**到一个地址连续的地址空间上，并以关键字在地址空间中的“象”作为相应记录在表中的**存储位置**，如此构造所得的查找表称之为**哈希表**。



地址空间存储的数据集合称为哈希表