



本节主题:

哈希文件

哈希表

哈希函数，散列函数

一个函数 $h(\text{key})$ ，用于根据关键字 key ，计算得到一个内存单元的地址 loc

关键字 $k_i (0 \leq i \leq n-1)$ 的存储地址： $h(k_i)$

哈希函数构造方法

直接定址法

除留余数法

数字分析法

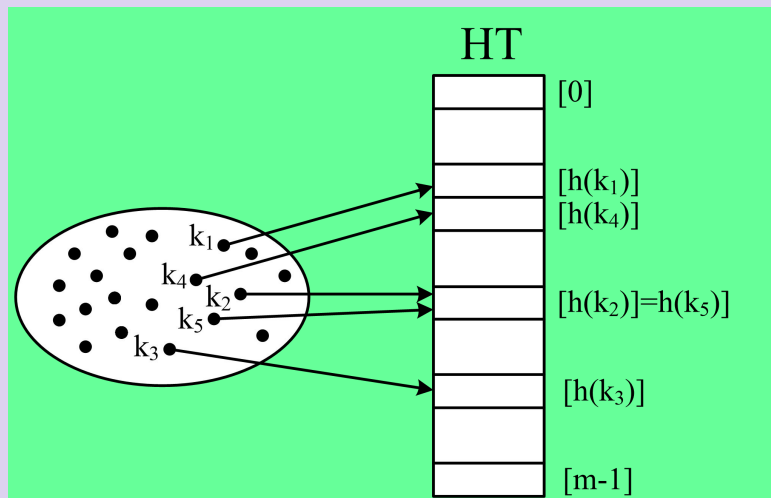
.....

冲突解决方法

开放定址法

拉链法

.....



哈希文件

❏ 哈希文件

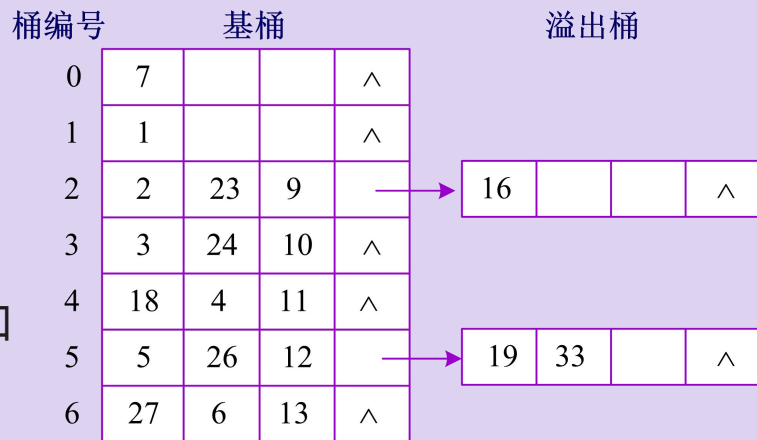
- ❏ 是利用哈希存储方式组织的文件
- ❏ 也称为散列文件、直接存取文件

❏ 与哈希表的相似处

- ❏ 根据文件中关键字的特点，设计一个哈希函数和处理冲突的方法，将记录哈希到存储设备上。

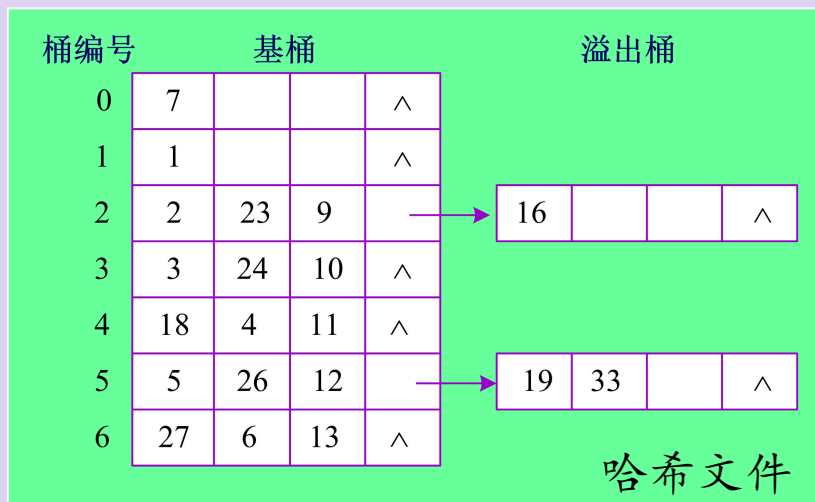
❏ 哈希文件的存储

- ❏ 磁盘上的文件记录通常是成组存放的，若干个记录组成一个存储单位，在哈希文件中，这个存储单位叫做“基桶”。
- ❏ 假如一个桶能存放 m 个记录，则当桶中已有 m 个同义词的记录时，存放第 $m+1$ 个同义词会发生“溢出”；溢出的数据存储到“溢出桶”。
- ❏ 处理溢出虽可采用哈希表中处理冲突的各种方法，但对哈希文件而言，主要采用链地址法。



溢出处理示例

- ❏ 某一文件有20个记录,其关键字集合为{2,23,5,26,1,3,24,18,27,12,7,9,4,19,6,16,33,11,10,13}。
- ❏ 桶的容量m=3，桶数b=7。
- ❏ 用除留余数法作哈希函数 $H(\text{key})=\text{key}\%7$ 。



基于哈希文件的操作

查找步骤

- 首先根据给定值求出哈希桶地址
- 将基桶的记录读入内存，进行顺序查找
 - 若找到关键字等于给定值的记录，则检索成功；
 - 否则，读入溢出桶的记录继续进行查找。

删除操作

- 在哈希文件中删去一个记录，仅需对被删记录作删除标记即可。

为了效率：希望同一哈希地址的溢出桶和基桶在磁盘上的物理位置不要相距太远，最好在同一柱面上。

桶编号	基桶				溢出桶			
0	7			^				
1	1			^				
2	2	23	9		16			^
3	3	24	10	^				
4	18	4	11	^				
5	5	26	12		19	33		^
6	27	6	13	^				

评价

❏ 哈希文件的优点

- ❏ 文件随机存放，记录不需进行排序；
- ❏ 插入、删除方便；
- ❏ 存取速度快；
- ❏ 不需要索引区，节省存储空间。

❏ 哈希文件的缺点：

- ❏ 不能进行顺序存取，只能按关键字随机存取
- ❏ 经过多次插入、删除后，也可能造成文件结构不合理，需要重新组织文件。

