

# 散列冲突处理

# 1 冲突

- 对于不同的关键字key1、key2，若 $key1 \neq key2$ ，但 $H(key1) = H(key2)$ 的现象叫做冲突。
- 具有相同的散列函数值的不同的关键字，称为该散列函数的同义词。

# 1 冲突

- 设计一个散列表时要考虑以下几点：
- (1) 确定散列表的地址空间范围，即确定散列函数的值域。
- (2) 构造合适的散列函数，使得对于所有可能的关键字，散列函数值都能在散列表的地址空间范围内，且出现冲突的可能性尽量地小。
- (3) 设计处理冲突的方法，即当冲突出现时该如何应对。

## 2 散列冲突处理

- (1) 开放定址法
- 当发生冲突时，形成一个探测序列，按此序列逐个探测散列表中的其他地址，直到找到一个开放地址为止，将发生冲突的记录放到该存储位置中。

## 2 散列冲突处理

- 散列地址的计算公式是：
- $H_i = (H(\text{key}) + d_i) \text{ MOD } m$ ,  $i=1, 2, \dots, k$  ( $k \leq m-1$ )
- 其中 $H(\text{key})$ 是函数， $m$ 是散列表长度， $d_i$ 是第 $i$ 次探测时的增量， $H_i$ 是第 $i$ 次探测时的散列地址。

## 2 散列冲突处理

- a. 线性探测法
- 将散列表 $T[0 \dots m-1]$ 看成头尾相连的一维数组。当发生冲突时，从该位置开始依次向后探测其他的地址，也就是说，增量序列为 $1, 2, 3, \dots, m-1$ ，直到找到一个开放地址为止。

## 2 散列冲突处理

- 优点：只要散列表未满，总能找到一个开放地址。
- 缺点：发生冲突的记录被散列到离冲突最近的开放地址上，从而增加了更多的冲突可能性(这种现象称为冲突的“聚集”)。

## 2 散列冲突处理

- 设关键字序列是26,7,30,8,40,34，散列表长度是7，散列函数是 $H(\text{key}) = \text{key} \bmod 7$ ，用线性探测法解决冲突，画出最后所得到的散列表。请大家课后做一下，看看结果是不是这样的。

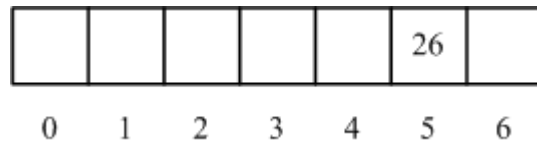


## 2 散列冲突处理

- 插入26前



- 插入26后



## 2 散列冲突处理

- 插入7前

					26	
0	1	2	3	4	5	6

- 插入7后

7					26	
0	1	2	3	4	5	6

## 2 散列冲突处理

- 插入30前

7					26	
0	1	2	3	4	5	6

- 插入30后

7		30			26	
0	1	2	3	4	5	6

## 2 散列冲突处理

- 插入8前

7		30			26	
0	1	2	3	4	5	6

- 插入8后

7	8	30			26	
0	1	2	3	4	5	6

## 2 散列冲突处理

- 插入40前

7	8	30			26	
0	1	2	3	4	5	6

- 插入40后

7	8	30			26	40
0	1	2	3	4	5	6

## 2 散列冲突处理

- 插入34前

7	8	30			26	40
0	1	2	3	4	5	6

- 插入34后

7	8	30	34		26	40
0	1	2	3	4	5	6

## 2 散列冲突处理

- b. 二次探测法
- 增量序列为 $1^2, -1^2, 2^2, -2^2, 3^2, -3^2, \dots, \pm k^2$   
( $k \leq \lfloor m/2 \rfloor$ )
- 优点：发生冲突的记录被跳跃式地散列到整个散列表中，不易产生冲突的“聚集”。
- 缺点：不能保证探测到散列表的所有地址。

## 2 散列冲突处理

- (2) 再散列法
- 构造一系列散列函数，当发生冲突时，利用不同的散列函数再计算下一个新的散列地址，直到找到一个开放地址为止。
- 优点：不易产生冲突的“聚集”。
- 缺点：增加计算时间。



## 2 散列冲突处理

- (3) 链地址法
- 将所有关键字为同义词的记录存储在同一个单链表中，并用一个一维数组来存储所有单链表的第一个结点的地址。
- 优点：不易产生冲突的“聚集”。

## 2 散列冲突处理

- 设关键字序列是26,7,30,8,40,34，散列表长度是7，散列函数是 $H(\text{key}) = \text{key} \bmod 7$ ，用链地址法解决冲突，画出最后所得到的散列表。请大家课后做一下，看看结果是不是这样的。

## 2 散列冲突处理

- 插入26前

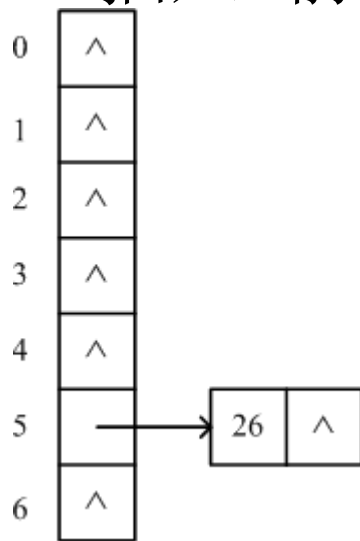
0	^
1	^
2	^
3	^
4	^
5	^
6	^

- 插入26后

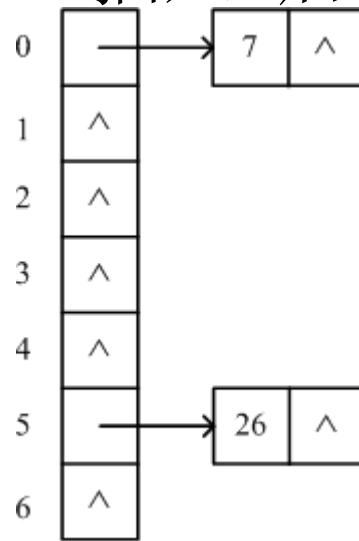
0	^
1	^
2	^
3	^
4	^
5	→ 26 ^
6	^

## 2 散列冲突处理

- 插入7前

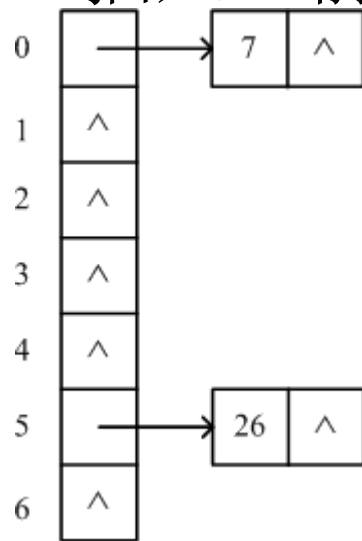


- 插入7后

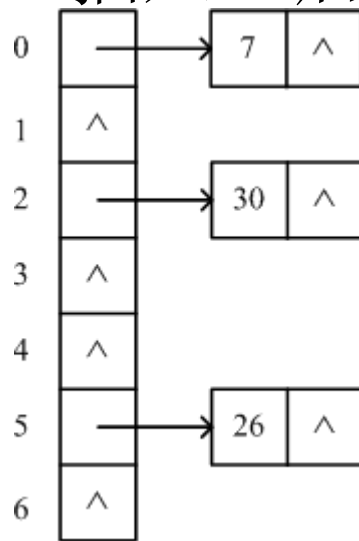


## 2 散列冲突处理

- 插入30前

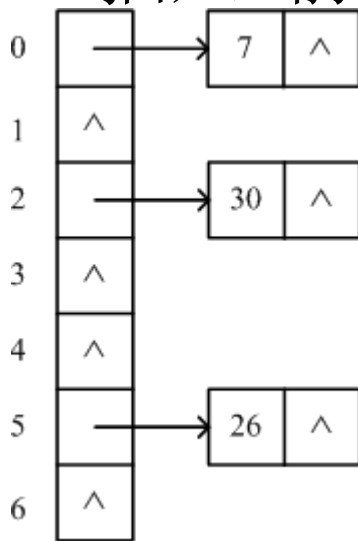


- 插入30后

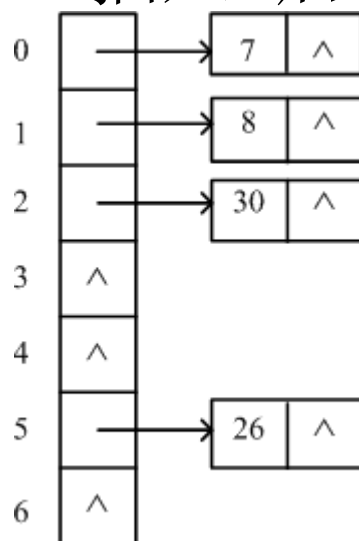


## 2 散列冲突处理

- 插入8前

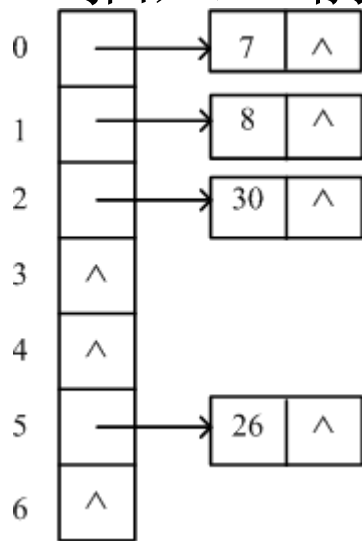


- 插入8后

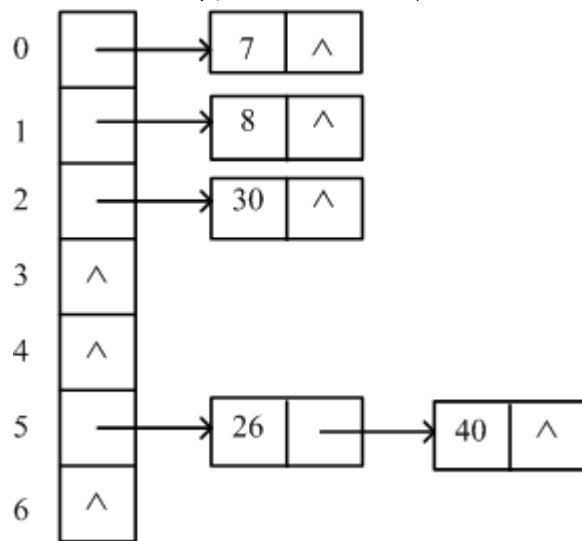


## 2 散列冲突处理

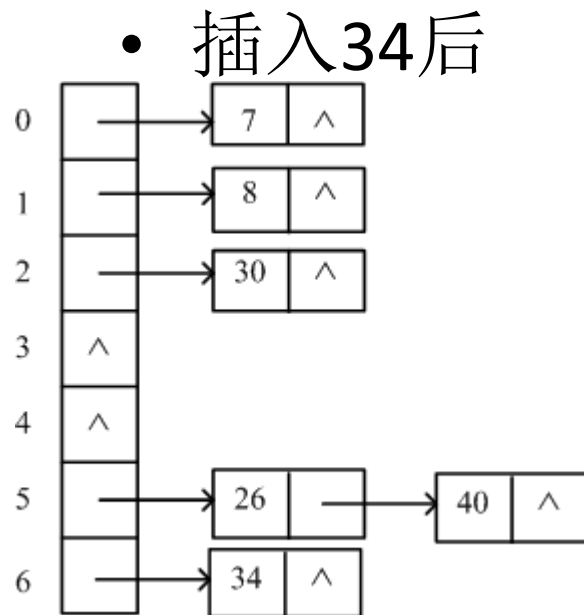
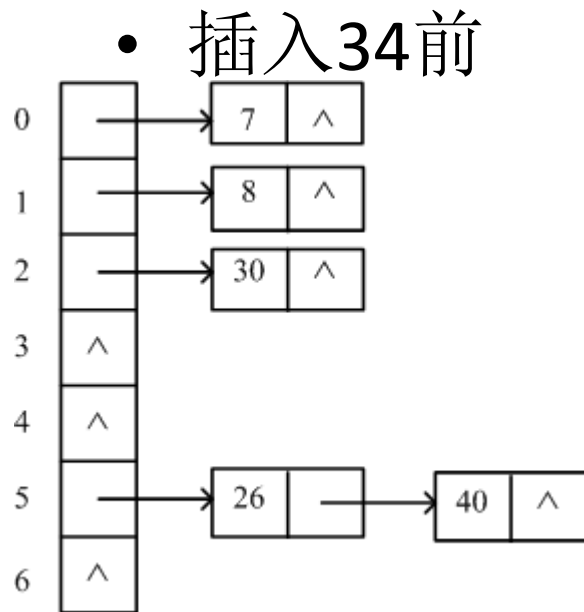
- 插入40前



- 插入40后



## 2 散列冲突处理





## 2 散列冲突处理

- (4) 建立公共溢出区法
- 在基本散列表之外，再设立一个溢出表，用来存储与基本表中记录冲突的所有记录。

# 思考

- 在散列表中如何删除记录？