



本节主题：

多关键字文件

由多样化的查询需求说起

□ 常有这样的问题

📁 学号为8的学生是谁？

📁 女生有多少人？

📁 有朝鲜族的同学吗？

📁 列出99102班的所有同学。

□ 次关键字索引及多关键字文件

📁 为了提高查找效率，还需要对可能会被查询到的次关键字建立相应的索引；

📁 这种包含有多个次关键字索引的文件，称为多关键字文件；

📁 次关键字索引可以是顺序表，也可以是树表 { 多重表文件
倒排文件

学号	姓名	性别	民族	班级
1	李明	男	汉族	99101
5	王平	男	汉族	99101
3	张萍	女	满族	99102
8	陈强	男	汉族	99101
4	马伟	男	回族	99102

为次关键字建立索引，
有必要！

多重表文件

多重表文件是将索引方法和链接方法相结合的一种组织方式。

例如，学生多重表文件

多重表文件对每个需要查询的次关键字建立一个索引，同时将具有相同次关键字的记录链接成一个链表，并将此链表的头指针、链表长度及次关键字，作为索引表的一个索引项。

物理地址	学号	姓名	性别	民族	班号	性别链	民族链	班号链
1	1	李明	男	汉族	99101	2	2	2
2	5	王平	男	汉族	99101	3	4	4
3	3	陈强	男	满族	99102	5	^	5
4	8	张萍	女	汉族	99101	^	^	^
5	4	马伟	男	回族	99102	^	^	^

(a) 多重表文件

次关键字	头指针	链长
男	1	4
女	4	1

(b) 性别索引

次关键字	头指针	链长
汉族	1	3
满族	3	1
回族	5	1

(c) 民族索引

次关键字	头指针	链长
99101	1	3
99102	3	2

(d) 班号索引

通常多重表文件的主文件是一个顺序文件。

多重表文件的操作

多重表文件的检索

先查询索引表

再在主文件中读出待查记录信息；

插入操作

插入时如果不要求保持链表的某种次序，则可将新记录插在链表的头指针之后；

删除记录

需要在每个次关键字的链表中删去该记录。

物理地址	学号	姓名	性别	民族	班号	性别链	民族链	班号链
1	1	李明	男	汉族	99101	2	2	2
2	5	王平	男	汉族	99101	3	4	4
3	3	陈强	男	满族	99102	5	^	5
4	8	张萍	女	汉族	99101	^	^	^
5	4	马伟	男	回族	99102	^	^	^

(a) 多重表文件

次关键字	头指针	链长
男	1	4
女	4	1

(b) 性别索引

次关键字	头指针	链长
汉族	1	3
满族	3	1
回族	5	1

(c) 民族索引

次关键字	头指针	链长
99101	1	3
99102	3	2

(d) 班号索引

倒排文件

- 在次关键字索引表中直接列出相关记录的物理地址或记录号，这样的索引表称为倒排表。
- 由主文件和倒排表共同组成倒排文件。
- 例：学生信息倒排文件
- 和多重表文件的区别

- 倒排文件中不必在多重表文件中增加数据，具有相同次关键字的记录不进行链接。

物理地址	学号	姓名	性别	民族	班号
1	1	李明	男	汉族	99101
2	5	王平	男	汉族	99101
3	3	陈强	男	满族	99102
4	8	张萍	女	汉族	99101
5	4	马伟	男	回族	99102

次关键字	物理地址
男	1, 2, 3, 5
女	4

(a) 性别倒排表

次关键字	物理地址
汉族	1, 2, 4
满族	3
回族	5

(b) 民族倒排表

次关键字	物理地址
99101	1, 2, 4
99102	3, 5

(c) 班号排表

评价

优点

- 检索记录较快，在处理复杂的多关键字查询时，可在倒排表中确定记录是哪个或哪些，继而直接读取之；

缺点

- 维护困难，在同一倒排表中，不同关键字的记录数不同，各倒排表的长度也不等。



百度一下