

按多关键字分类与索引的异同

周 瑛 (安徽大学信息管理系)

摘要: 本文通过对 dBASE III 中的分类和索引操作的详尽分析, 细致地阐述了这两种重要操作结果的异同, 并指出了目前教科书和参考书中普遍存在的问题。

在 dBASE III 中, 经常用到分类和索引这两种操作来形成一个有序的库文件, 便于提高查询速度, 节约查询时间。

分类是按照关键字的大小来重新排列数据库中的记录, 其格式为: SORT TO<新文件名>ON<字段 1>[/A] [/D][<字段 2>[/A] [/D]...] [范围][FOR<条件>]。而索引的建立, 则是按关键字的升序对当前数据库中的记录进行排序, 其格式为: Index on<关键字 1>[+<关键字 2>[+...]] TO<索引文件名>。这两条语句中的关键字除了不能是逻辑型和记忆型的字段外, 可以为字符型、数字型或日期型字段。只按一个关键字进行分类和索引, 若仅从这两种操作的执行结果来看, 它们所实现的功能是相同的(按升序排列), 但若按两个或两个以上(即多个)关键字来进行分类和索引, 则要根据关键字的类型来看它们的执行结果, 既有相同之处, 又有差别, 下面以两个关键字为例加以说明。

设现有一学生数据库 xs.dbf, 该库包含五个字段, 库结构显示如下:

```
Structure for database:C:\WPS\XS.DBF
Number of data records: 6
Date of last update:04/08/94
Field Field Name Type Width Dec
1学号 Character 6
2姓名 Character 6
3性别 Character 2
4数学 Numeric 3
5英语 Numeric 3
* * Totas * * 21
```

该数据库的内容显示如下:

Record#	学号	姓名	性别	数学	英语
1	100001	王民	男	95	50

2	100002	田国	男	90	60
3	100003	吴月	女	80	80
4	100004	孙英	女	85	95
5	100005	李辉	男	85	60
6	100006	李薇	女	85	85

首先, 对字符型字段学号、姓名进行分类, 操作步骤及结果显示如 F:

```
.use xs
.SORT ON 学号,姓名 TO XXXS
.use xxxs
.LIST
Record# 学号 姓名 性别 数学 英语
1 100001 王民 男 95 50
2 100002 田国 男 90 60
3 100003 吴月 女 80 80
4 100004 孙英 女 85 95
5 100005 李辉 男 85 60
6 100006 李薇 女 85 85
```

再对字符型字段学号、姓名进行索引, 操作步骤及结果显示如下:

```
.use xs
.INDEX ON 学号+姓名 TO XXS
.List
Record# 学号 姓名 性别 数学 英语
1 100001 王民 男 95 50
2 100002 田国 男 90 60
3 100003 吴月 女 80 80
4 100004 孙英 女 85 95
5 100005 李辉 男 85 60
```

然后, 对数字型字段数学、英语进行分类操作, 步骤及结果显示如下:

```
.use xs
.SORT ON 数学,英语 TO YXS
.use yxs
```

.list

Record#	学号	姓名	性别	数学	英语
1	100003	吴月	女	80	80
2	100005	李辉	男	85	60
3	100006	李薇	女	85	85
4	100004	孙英	女	85	95
5	100002	田国	男	90	60
6	100001	王民	男	95	50

再对数字型字段数字,英语进行索引,操作步骤及结果显示如下:

.use xs

.INDEX ON 数学+英语 TO SXS

.list

Record#	学号	姓名	性别	数学	英语
1	100001	王民	男	95	50
5	100005	李辉	男	85	60
2	100002	田国	男	90	60
3	100003	吴月	女	80	80
6	100006	李薇	女	85	85
4	100004	孙英	女	85	95

从上述操作及结果可以看出,按多关键字进行分类或索引操作时,若关键字为字符型字段,则两操作的结果完全相同,即先按第一个关键字进行分类或索引,第一个关键字值相同时,再按第二个关键字进行分类或索引,依此类推。若关键字为数字型字段时,两操作的结果迥然不同。对分类操作来说,仍然和字符型关键字一样,先按第一个关键字进行排序,第一个关键字值相同时,再按第二个关键字进行排序,如此等等。但对索引操作来说,则与字符型关键字不一样,该操作是按几个数字型字段值的总和来排序的,总和相同的记录,原数据库中排在前面

的记录,在索引文件中仍排在前面,也就是说,首先是按照数字型关键字段值的总和建立索引,因此,显示的执行结果是按照总和从小到大的顺序来排列的,对总和相同的,按原来顺序来排列,不再重排(见上例,在索引文件中,记录按照数学和英语两值之和从小到大的顺序排列,若总和相同,象记录1和记录5,则按原库文件xs.dbf的顺序排列,因此,记录1排在记录5的前面,而不象分类操作那样,记录是按照数学与数从小到大排列,数学分数相同的,象原库文件中的记录4、5、6,则再按英语分数从小到大排列,这一点,是索引操作和分类操作结果的主要区别。我们在实践中,要认真区别对待,不能象参考书中所提的那样,把实际不同的两种操作结果等同起来。对索引操作来说,在多个数字型关键字的情况下,要得到与分类操作相同的结果,用STR函数,先将数字型字段转换成字符串,再建立索引,即可得到与分类操作结果相同的索引文件。

参考书:

[1]毛玉皎主编《图书情报管理自动化基础》武汉大学出版社1993年1月第一版

[2]郑阿奇编《汉字dBASEⅢ(IV)实用及程序设计技巧(含FOXBASE+)》东南大学出版社1992年5月第二版

[3]王治宇主编《dBASEⅢ程序设计与实用技术》电子工业出版社1991年3月第一版

[4]王云宜等编《汉字dBASEⅢ原理与应用》(修订版)电子工业出版社1990年10月第一版