

xBASE

eBASE

数据库

计算机科学1996 Vol. 23 No5

54-56.78

xBASE 家族的新成员——eBASE^{*})

周宗毅 李磊

TP311.13

TP311.13

(中山大学岭南(大学)学院软件研究所信息科学实验室 广州510275)

摘要 This paper gives the description of eBASE, a fully open, high-speed multimedia database management system. This system is compatible with the former system of xBASE, efficient in time and space, portable and easy to use. eBASE has been applied in several application development projects, and it is proved that eBASE can be embedded in different application systems as practical database engine.

关键词 Relational database management system, C language, Function library.

人们已经不再满足于对数据的简单管理、查询、统计等,还希望用已有的数据对未来的情况进行预测、决策,提出建议(专家系统);另一方面,人们希望数据库管理系统不但能够处理传统的可以用代码表示的数据,而且能够处理图象、声音等信息。因此,目前数据库技术正处于突变的时期。新一代的数据库管理系统应当具有这样的特性:①良好的继承性;②全开放性;③良好的时间效率;④良好空间效率;⑤多媒体的能力;⑥用户接口简单方便。

eBASE的发展经历了 eBASE I 和 eBASE II 两个阶段。eBASE I 是小巧精致的数据库引擎,主要实现逻辑查询和多媒体信息存取。eBASE II 具有全面的数据操纵能力,已经经过应用的考验。

1 eBASE 的特性

1.1 属于 xBASE 家族

在商品化的微机关系数据库管理系统中,以 dBASE 为代表的 xBASE 系列尤其引人注目。eBASE 同 xBASE 的数据库产品完全兼容。这包括三个方面:①支持 xBASE 数据库管理系统的功能,保持以元组为处理对象的主要特征。②数据文件格式采用已事实上成为工业标准的 dBASE III 的格式,能够使用以往所有积累起来的数据。③在多用户环境下,使用与 xBASE 相同的加锁协议,可以实现与 xBASE 实时的数据共享。

eBASE 的函数功能与函数命名风格与 xBASE 接近。这样,用户可以方便地将开发平台由 xBASE 转到 eBASE 上。

1.2 全开放

xBASE 在操作系统之上建立自己的环境,是应用程序的功能以及实现技术的依托。但是,随着计算机应用的深入,应用程序并不仅仅面向某一单纯的应用,尤其在目前多媒体技术不断发展的时候,有多媒体技术应用的地方,就有数据操纵问题,就有必要使用数据库技术。eBASE 以 C 语言函数库的形式提交给应用系统的开发者,全部数据库语句以 C 语言函数的形式嵌入在 C 语言程序中使用,或者在任何具有 C 语言接口的开发平台上使用,例如 Visual Basic 和 AutoCAD, eBASE 与宿主语言的结合是无缝的,不存在耦合的代价。

1.3 增加新的字段类型

eBASE 与 dBASE III 相比增加了 Image 类型的字段和对应于该类型字段的存取函数。在应用实践中,这一类型字段不仅能用来存取图象信息,还能够存取所有二进制数据类型的信息,也就是说,可以存取多媒体信息。eBASE 与其它媒体获取与播放子系统一起,为用户提供开发多媒体应用程序的手段,减轻了开发难度。

1.4 逻辑查询功能

逻辑查询是关系数据库管理系统必不可少的功

*) 本研究由广东省自然科学基金和中山大学重点科研基金资助。周宗毅 硕士生,李磊 教授。

12

能,eBASE 的逻辑查询有以下特性:①有一套完备的逻辑运算符,构造的查询表达式的表达能力没有任何限制;②有较高的时间效率.在 eBASE 中,逻辑查询表达式的 BNF 定义为:

〈逻辑查询表达式〉::=(〈逻辑查询子表达式〉{ AND, 〈逻辑查询子表达式〉}
 〈逻辑查询子表达式〉::=(〈属性关系表达式〉{ OR, 〈属性关系表达式〉}
 〈属性关系表达式〉::=(〈字段名〉〈关系运算符〉〈常量〉
 〈属性关系表达式〉::=(〈常量〉〈关系运算符〉〈字段名〉
 〈属性关系表达式〉::=(〈字段名〉〈关系运算符〉〈字段名〉

并且规定运算顺序为:关系运算符、“或”逻辑运算符、“与”逻辑运算符,而且同级运算符从左到右计算,根据以上定义,逻辑查询表达式中不使用“非”逻辑运算符,不过,由于 eBASE 所定义的关系运算符都具有对称性,这种规定只限制表达形式而没有限制表达能力,另外,逻辑查询表达式之所以使用这一范式形式,是因为:①任何一个逻辑表达式都可以转化为这样的范式形式,②与组合查询的界面一致,方便程序自动生成查询表达式,③方便解释,④是非逆波兰的表达式计算方法,可依靠逻辑学的成果进行优化.

eBASE 对逻辑查询表达式的关系运算符集进行了改善,可增强表达能力.其一是字符型字段的模板匹配运算,即允许查找具有某一特征的属性值,它非常适合进行模糊查找的需要,其二是对日期型字段原子成分的处理,eBASE 允许在日期常量中加上标记,使得关系运算符的处理对象由日期降低为对日期的某些成分的处理.

eBASE 除了采用逻辑学成果对逻辑查询进行优化外,还使用了缓冲区技术和 C 语言的编码优化技术,因此,eBASE 获得了较高的逻辑查询速度.[查询时间比约为 dBASE],Clipper 5.01:FoxBASE+2.1:FoxPro1.0:eBASE I=5.25:3.75:2.10:1.50:1.00]

2 eBASE 的具体实现

2.1 eBASE 函数库

eBASE 提交给用户的是一系列 C 语言头文件、函数库,以及 Windows 下的动态连接库,用户可根据不同的应用环境——无论是单用户还是多用户,无论 DOS 还是 Windows——使用合适的函数库,如同使用 C 的标准库函数一样进行编程.

eBASE 的库函数包括:

①系统管理函数,实现功能包括系统的初始化和退出.

①数据库管理函数,实现功能包括数据库文件的创建、打开、关闭、整理、清空和排序.

②记录管理函数,实现功能包括记录(也包括单个字段)的增加、删除、修改和读取,

③记录指针操作函数,实现功能包括记录指针的绝对移动和相对移动.

④逻辑查询函数,实现功能包括表达式的翻译和查找.

⑤索引管理函数,实现功能包括建立索引、重建索引、去除索引、选择主索引、按索引查找等.

⑥数据库信息函数,提供从字段、记录到整个数据库文件提取信息的功能.

⑦共享数据管理函数,多用户环境下的特定功能穿插以上各类函数中,这一类函数主要提供实现加锁协议和缓冲区刷新的函数.

⑧静态索引管理函数,其中一个静态索引函数根据指定的数据库文件、关键字(可多至16个字段)和逻辑查询表达式,从原数据库文件中找出符合条件的记录,按关键字排序,并将这些记录的记录号和组成关键字的字段按排序结果存入新建的数据库文件(结构由关键字决定).新库与原库之间可通过另外的函数进行关联操作.静态索引函数对数据库同时进行投影和选择计算,是为了适应应用中对集合操作的需求.

⑨实用函数,包括日期表示的转换、数值表示的转换、字符串处理函数,

eBASE 的库函数具有以下特性:①库函数精简,能实现全部功能.用户需要掌握的函数不多,易于上手,②良好的封装性,eBASE 将内部数据结构的内部函数封装起来,并不公开给程序,所公开的数据结构只有日期内部数据结构一种,所公开的编程接口函数都带有安全检查机制,避免非法参数造成破坏.

2.2 eBASE 的字段类型

类型	标识符	长度(字节)
字符	C	1~255
数值	N	1~255
日期	D	8
逻辑	L	1
备注	M	任意
图象(二进制)	I	任意

2.3 eBASE 的数据文件类型

①数据库文件(.DBF),用于存放数据库的结构和记录以及每一个字段的数据——备注和二进制字段例外,因为它们的信息存储在辅助文件中,当不含

二进制字段时,文件格式与dBASE III完全兼容。

②备注字段文件(.DBT),存放某一数据库文件的所有备注字段。其文件名与所属的数据库文件相同,但扩展名不同。文件格式与dBASE III完全兼容。

③图象(二进制)字段文件(.DBI),存放某一数据库文件所有的二进制字段。命名规则同②,文件格式为eBASE独有。

④索引文件(.BNX),eBASE允许对同一个数据库按不同的字段索引,所有的索引信息存放在同一个文件中,即一个数据库文件至多有一个索引文件。命名规则同②,文件格式为eBASE独有。

2.4 配套工具

与eBASE相配套的还有图形界面函数库EU-IMS及一些数据库实用工具,如EKIT。

EKIT是eBASE维护数据资源的集成实用工具,并且在数据库可视化方面作出一些尝试。此外,由于eBASE的应用系统主要在C语言平台上开发,为使编码规范化,增加程序的可读性、可维护性,EKIT担负起联系数据库和源程序的桥梁作用。

EKIT一方面能根据数据库描述表文件(一种可作为开发文档的文本文件),生成空的数据库文件和辅助文件,并且生成描述数据库信息的C语言头文件和源程序,在这些文件中定义了描述数据库各方面信息的宏名字,数据结构、结构数组等等。另一方面,能从数据库文件反过来生成数据库描述表文件。

3 性能比较

CodeBASE是由美国Sequiter软件公司开发的较为成熟的数据库C函数库,是与eBASE同类的数据库引擎。在这里对eBASE I(多用户版)与CodeBASE 5.0所作的比较包括速度、资源占用情况等方面。

3.1 速度的比较

测试1 读一个数据库的1000条记录并显示第1个字段的内容。结果:eBASE读记录的速度较快,因为eBASE I程序中定义了与数据库内部格式基本一致的记录缓冲区,所有类型字段按字符方式处理,与文件存储格式一致,而CodeBASE 5.0在读取过程中无论是否需要都要做类型转换。

测试2 向同一个数据库追加1000条记录。结果:eBASE I追加记录速度较快,这跟eBASE I操作比CodeBASE 5.0简单、直接有关。

测试3 对同一数据库的同一条记录连续修改1000次。结果:eBASE I修改速度较快,原因同上。

测试4 使用索引查找同一数据库的同一记录。结果:CodeBASE 5.0索引查找速度较快,原因是多用户版的eBASE I对索引数据要做更多安全保护工作。

测试5 用相同的逻辑表达式查询同一数据库。结果:eBASE I的逻辑查询速度明显比CodeBASE 5.0快2至3倍。

测试6 网络环境下对同一数据库追加记录,以测试多用户环境对速度的影响。结果:CodeBASE 5.0受多用户环境影响更大些。

测试7 网络环境下同时对同一数据库作1000次相同的修改。结果,同上。

测试8 网络环境下同时对同一数据库用相同的逻辑表达式做逻辑查询。结果:在网络环境下,并行查询时eBASE I的逻辑查询速度会降低,但程度不大,而CodeBASE 5.0的速度就大为降低。原因是eBASE I在查询时不需要给文件或记录加锁,而CodeBASE 5.0的逻辑查询功能是用关联实现的,需对文件加锁。

3.2 内存占用情况的比较

分别使用eBASE I和CodeBASE 5.0编写一个打开数据库的程序,在运行过程中观察内存使用情况,发现:eBASE I占用内存远少于CodeBASE 5.0。实际上,eBASE I允许用户平衡效率和空间的需求,调整数据库占用的内存数目,少至1K以下。在复杂的应用系统中,尤其是多媒体系统,基本内存的使用必然紧张,因而,eBASE一方面减少基本内存的占用,另一方面使用XMS或EMS进行排序等数据量大的操作。

3.3 外存占用情况的比较

考虑3.2所用的两个可执行文件发现:eBASE I占用的外存资源比CodeBASE 5.0少。

3.4 缓冲区大小变动对速度的影响

在打开数据库文件时,eBASE I允许指定记录缓冲区的大小。现通过变动测试5的缓冲区大小来观察对逻辑查询速度的影响。一般而言,eBASE I在一定范围内,缓冲区越大,逻辑查询速度越快,并且经验表明,对越大的库和越复杂的逻辑表达式,影响越明显。CodeBASE 5.0的记录缓冲区对程序员透明,因此程序员无法更改缓冲区大小。

结束语 eBASE在Borland C++平台上,全部使用C语言及标准库函数编程实现,代码移植方便,既可以

(下转第78页)

式定义交由模式处理模块生成 IGES 字典文件 (IGES. afd) 和 IGES 早联编库程序实现文件 (IGES. cpp), 在 STEP. cpp 和 IGES. cpp 中的类定义中各自封装了相应的映射函数原型, 映射函数的程序实现文件 (I-Smap. cpp) 是由映射处理模块产生的 (用户可在其中加入复杂的映射算法)。

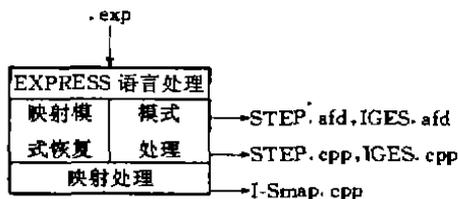


图 5 ToMap 功能结构

STEP/IGES 早联编类的数据结构如图 6 所示, 它们在原有 ASWF++ 早联编类数据结构的基础上增加一个数据成员 MapObj 和两个成员函数 MapCreate 和 MapRelation, MapObj 为指针变量, 用来指向其映射对象, 当执行 MapCreate 操作时对其赋值 (包括映射对象的 MapObj 值)。MapCreate 和 MapRelation 均为映射函数, 当执行 MapCreate 时则创建了映射对象, 并对映射对象的非引用属性 (基本属性) 进行赋值, 在执行 MapRelation 后才对引用属性进行赋值, 即恢复对象之间的引用关系。因此在映射控制器中进行映射转换时, 实际上分了两步骤, 即首先执行所有对象的 MapCreate 操作, 然后再执行所有对象的 MapRelation 操作。为了能够在迟联编

库中执行这些操作, MapCreate 和 MapRelation 均为虚函数。

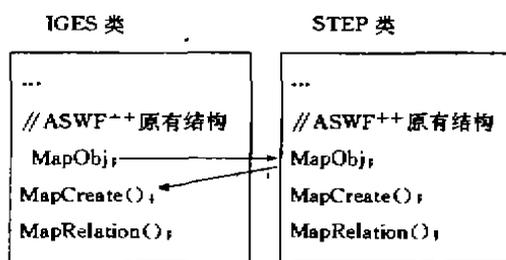


图 6 STEP/IGES 类数据结构

结论 ASWF++ 是面向对象的 STEP 数据管理软件, 它提供的数据映射机制可以大大简化 IGES-STEP 转换器的开发工作, 这项研究工作也同样适合于其它类型的 STEP 数据转换器。如开发 DXF-STEP 转换器, 这时只要对图 4 中的. exp 文件模式定义及映射信息进行修改, 并开发相应的数据文件读写器即可。

参考文献

[1]ISO CD 10303-22,1993(SDAI)
 [2]ISO IS 10303-11,1993(EXPRESS)
 [3]Rose Library Reference Manual,STEP Tools, Inc.
 [4]Dou Hemmelgarn, IGES to STEP Migration, AutoFact' 91
 [5]IGES V4.0,NBSIR88-3813

(上接第 56 页)

做成 MS-DOS 环境下的函数库, 又可以做成 Windows 环境下的动态连接库。更重要的是, eBASE 内核精致小巧, 但具有高效、实用的特点, 它可以作为数据库引擎嵌入到不同的应用系统中, 便于修改、扩充, 具有相当的灵活性和适应性。

目前, eBASE 已经达到实用水平。eBASE I 被用来为广州市信息中心开发了广东省产业信息咨询系统, 并受到用户高度赞扬。eBASE I 更被用来开发多个多媒体演示系统及多媒体信息管理系统, 其中与广州空军后勤部航材处合作开发的“军区微机航材信息管理系统”是一个大型的项目, 仅数据库文件就有 480 多个, 曾在空军系统的应用软件评审中

获奖。香港东亚安泰保险公司原来使用 Visual Basic 与 Access 开发的人身保险辅助销售系统的计算时间是分钟级的, 后来改用 eBASE I 重新开发, 使得计算时间降为秒级。

参考文献

[1]萨师焯等, 数据库系统概论. 高等教育出版社, 1983
 [2]严蔚敏等, 数据结构. 清华大学出版社, 1987
 [3]周龙骧, 数据库管理系统实现技术. 中国地质大学出版社, 1990
 [4]S. Ceri, G. Gottlob and L. Tanca, Logic Programming and Databases, Springer-Verlag, 1990