

扩充数据项图法与微机关系数据库概念结构设计

长沙水电师范学院计算中心 柳见成

摘要: 本文通过实例分析了应用扩充数据项图法进行数据库概念结构设计的过程。认为与当前流行的 E-R 图法相比,扩充数据项图法更适于一般微机用户的需要,同时讨论了在微机关系数据库设计中应用这一方法的若干特点。

一、引言

数据库结构设计是管理信息系统设计中极为重要的一个环节,结构设计的质量直接影响到数据库的使用效率和整个系统的生命周期。目前进行数据库结构设计主要采用 SA/SD 方法,这是一种以数据分析为核心,结合功能分析的结构化系统分析方法。由于它具有较成熟的规则和工具(如数据流图和数据字典等)可供使用。能较合理地获得符合 ANSI 标准的三级数据库模式结构,并借助于 SD 方法设计出操纵数据库所需的应用程序,因此已成为数据库结构设计中的主流方法。然而另一方面,由于 SA/SD 方法需要根据系统的功能分析确定 E-R 图,将 E-R 图转化为不可化简单元集,然后导出相应的有向图和连接矩阵,并对其施行传递闭包算法和最小覆盖算法以获得数据库系统的概念结构[1]。这一过程的复杂性和数学上的难度,对于不具备图论和关系代数理论知识同也没有系统学习过数据库设计的一般微机用户来说,是一个难以克服的困难。正因如此,目前介绍微机关系数据库的书籍通常只对 E-R 图法作简单的举例(或根本不作介绍),而熟悉 dBASE/FOXBASE 系统的应用人员在实际中也不应用这一方法,而是凭经验进行试凑法设计。这就造成了一方面 E-R 图法难以推广,另一方面广大微机用户在应用数据库设计中又缺乏一种系统的实用方法。

扩充数据项图法是 H. C. Smith 于 1985 年提出来的[2]。它不必建立 E-R 模型,在方法上也不以功能分析为主,而是要求用户在数据需求分析阶段详细了解系统的各数据项信息及它们之间的依赖关系,并据此作出依赖表,然后直接根据扩充数据项图法规则作图,导出完

全规范化的关系表,从而比较直观地获得系统概念模型。本文结合一个简化的教学活动管理模型,介绍应用该方法获得概念结构模型的步骤,并分析其在微机关系数据库设计中的若干特点。

二、应用扩充数据项图法设计数据库概念结构的过程

1. 数据需求分析

在数据需求分析阶段,不仅应考虑到信息系统的当前需求,也应考虑到今后一段时间内可能提出的新需求。需求分析应当产生的系统目标、需求定义、系统总体结构及功能说明等内容,对于各种设计方法都是基本相同的。扩充数据项图法的特点在于直接面向数据和数据项间的联系,以及数据项间的结构分析。因此应详细了解系统所涉及到的数据项信息(数据项名、数据项类型、数据项的字节长度及有效取值范围等)以及数据项之间的依赖关系。依赖是指数据项之间的关联,这里只考虑 1 型关联和复杂关联,对于条件关联则看作是复杂关联中的一种情况。

数据需求分析通过描述数据项之间关联的“依赖语言”来构成“依赖表”,用以记录该系统的结构信息,描述“依赖语言”主要有两条原则:(1)对涉及到的每个数据项分配一个符合所使用的 DBMS 软件规则的名字。例如在 dBASE/FOXBASE 系统中每个字段最多允许 10 个字符或 5 个汉字作名字。(2)“依赖语言”主要描述 1 型关联(函数依赖)或复杂关联(多值依赖)。

例如,分析一个学校的教学管理系统后,可获得如下依赖表:

①一个系有多个班级和多个教师；②每个班级有多种教科书；③每种教科书有一个编号；④每个班级有多个学生；⑤一周内每个班级有多个课时；⑥在一个给定课时内在一个教室上课，每个课时内上一门课程，且由一名教师上课；⑦每个班级修多门课程；⑧每个学生有一个性别、一个学号和年级；⑨每个教师可教多门课程。

还应当指出，对于扩充数据项图来说，尽管主要关心的是数据项及其之间的联系，但要最终获得高质量的关系模式，仍应进一步了解有关数据操纵方面的信息。如某一数据项是仅被检索，还是需要进行频繁的存取更新操作等，这样才能确定包含这些数据项的关系到底应当规范化到什么程度为好。通常是通过建立数据字典来进行说明。建立数据字典对熟悉 dBASE / FOXBASE 系统的应用者并不困难(尽管 dBASE / FOXBASE 系统没有这方面的专用功能)，但这却是易忽略的问题。

2. 概念结构设计

数据库概念结构设计是根据用户需求和处理要求形成系统的概念模型。扩充数据项图法在概念结构设计阶段的特点是不从实体及其联系出发，而是从作为基本属性的数据项及数据项之间的依赖关系出发，综合出实体及实体间的联系，扩充数据项图法建立在用图来描述数据项之间的单值函数依赖和多值依赖的基础上。下面介绍扩充数据项图法的一些基本术语。

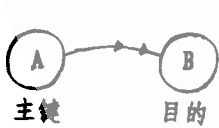


图1 单值依赖



图2 多值依赖

扩充数据项图中的每个数据项用圆圈包围，称为泡；数据项之间的关联(依赖)用箭头表示，称为键。图1是单值依赖，A称为主键泡，B称为目的泡。图2是多值依赖，C称为主键泡，D称为尾键泡，把C→→D称为一个尾键依赖。

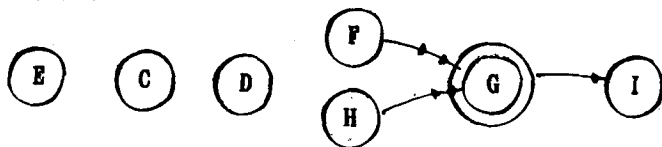


图3 多值依赖中的上链键

图4 用双泡表示不同的联接含义

当另有一个双箭头指出主键泡时，如图3，则把E称为一个上链键。如果一个泡在图中具有双重身分，如图4中，F→→G, H→→G→→I，为了区别可将G用双泡表示。

借助于扩充数据项图进行概念结构设计包括作图、获得规范化关系表及综合得到概念结构模型三个基本过程：

(1) 根据数据需求阶段产生的“依赖表”作出“数据项依赖图”。作图开始前应先确定核心数据项。核心数据项既不是表达多个数据项之间的联系，也不是用来限制或描述其它数据项。如教学管理系统中的“系”就属于这样的数据项。在数据依赖图中每个数据项只能出现一次，如有必要可根据依赖表修改局部数据项名。

作图时从核心数据项出发，逐步用依赖表中的依赖语句扩大依赖图，并在依赖图的连结上依次标上依赖语句编号，以检查是否有依赖语句被遗漏掉。

根据上例依赖表作出的数据项依赖图如图5所示。

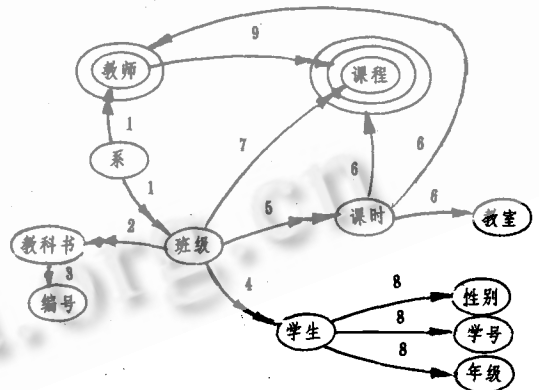


图5 教学管理系统数据项依赖图

(2) 将扩充数据项图转化为规范化关系表

在单值依赖和多值依赖连接的图中，将图转化为规范化关系表的原则是：将上链键与主键一起构成表的主键。此外按两种基本情况讨论。

①在泡所形成的链(称为泡链)中，如果没有目的泡，就将尾键、主键和上链键合并成一张全键表。如图6中虚线给出的表B1就是一张全键表。下划线表示该数据项为键。

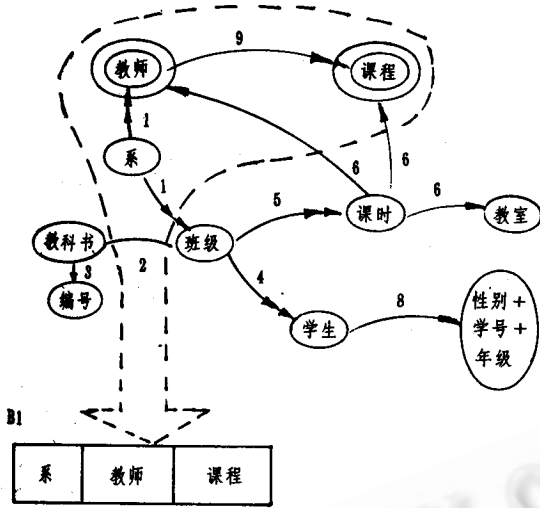


图6 没有目的泡的泡链合并成全键表

②在泡链中,如果有目的泡,则沿指向目的泡的箭头逆向跟踪到最后一个上链键泡。这种泡链中所有上链键和主键构成关系表的主键,所有的目的泡构成表的一般属性。如图7中虚线给出的表B2中包含三个主键和三个普通属性数据项。

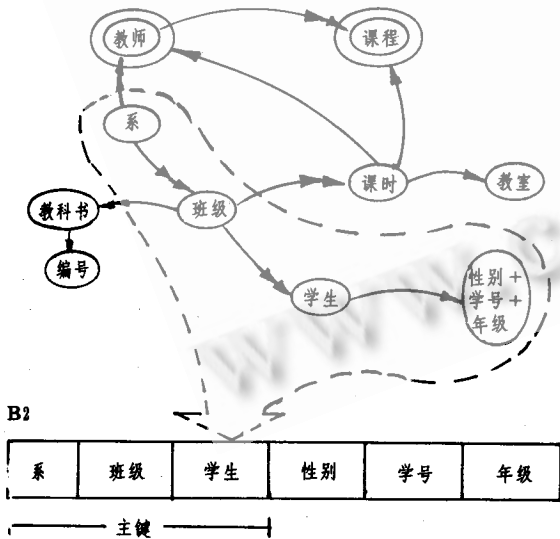


图7 含有目的泡的数据项依赖图转化为关系表

按同样规则还可得到表B3和B4:

B3

系	班级	课时	教师	课程	教室
---	----	----	----	----	----

B4

班级	教科书	编号
----	-----	----

本例中没有讨论孤立泡和当合并所得表中主键超过三个时,借助于一个“临时”中间泡修改数据项图,将原表分解为含有不超过三个主键的两个关系表的特殊情况。

(3)将上述过程中求得的关系表综合成关系数据库概念模型,所有外部键用线连接起来。如图8所示。

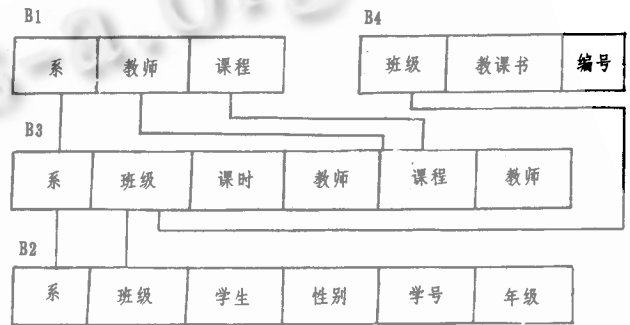


图8 对应依赖图5的关系概念结构模型

规范化关系中各范式对数据项依赖图的要求如下:

1 NF: 如在转换时保持每一数据项是不可分的(原子的),则转换到关系表中也是不可分的。

2 NF: 要求每个非键数据项都完全依赖于主键,显然这一点可以从泡链上直观地得到。

3 NF: 要求依赖图中删去传递依赖,这相当于转换成关系时,通过分解消除传递依赖性。如图5中“班级”、“课时”、“课程”构成了传递依赖,如图9所示,故在图6中删去了依赖线7。

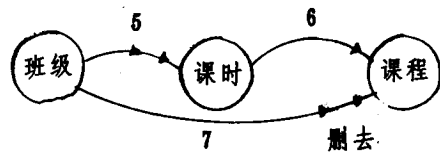


图9 图5中的传递依赖

4 NF: 在3 NF的关系表中不能包含两个以上的多值依赖,如出现则应分解掉。如在图10中应将其分散为

两个全键关系表。

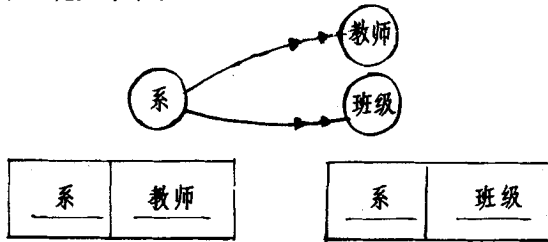


图 10 分解两个以上的多值依赖

三、微机关系数据库设计中应用扩充数据项图法的特点

微机关系数据库用户一般对用户视图比较了解,即熟悉系统涉及到的各种数据项及其之间的依赖关系,知道当前的需求和将来可想象到的需求,同时熟悉 dBASE / FOXBASE 系统的使用。一旦获得概念模型并将其转化为数据库结构,则对于各种数据操作如插入、删除、输入、输出、编辑和查询等操作的处理基本形式就已形成。因此,通常将关系表转化为库结构后即可着手编程。

dBASE / FOXBASE 系统中不支持对三级模式的直接描述,模式与用户子模式基本上是无区分的,数据库结构的模式也就是它提供给各用户的子模式。完成数据库结构设计只需一次映射。这样虽然降低了数据的独立性和安全性,并且增加了用户接口的复杂程度,但另一方面也带来了使用方便,容易掌握的特点。不难看出,扩充数据项图法是比较适合于这种不区分模式与子模式的 ROBMS,用手工完成设计的场合的。一旦获得概念结构模型;在逻辑结构设计时就可直接用 DBMS 提供的关系定义语句(如 FOXBASE+中的建库命令 CREATE 和索引命令 INDEX ON)进行描述。

设计者在数据需求分析阶段获得明确和详细的需求定义,即收集与系统业务有关的一切数据项的信息,并通过编号的数据依赖语句整理出该系统的“依赖表”。由于扩充数据项图法不是面向数据流和功能分析,也不产生 DBF 图,一切原始数据项及其关联都通过依赖表来描述,因此这一阶段应特别仔细。根据依赖表作图除遵循基本规则外,还应注意避免依赖关系过于含糊和复杂。如当多个多值依赖指向同一泡时使用多泡表示明确语义,能用单值依赖表达的就不要用多值依赖表达等(4)。

从依赖图转换到关系表的关键,在于正确无遗漏地搜索依赖图中所有链泡键。

微机数据库用户常遇到的问题是程序难以编写和维护。如信息冗余太大,数据库文件太多,查询速度太慢等,往往需要通过修改原数据库结构来提高运行性能。如信息冗余过大,可能是库结构中某些字段的值重复性太大,占用的字节空间过多,从而影响查询速度。这时应对原库结构进行分解。又如当程序频繁交叉访问一些具有不同数据项的库时,会大大增加系统的开销。从概念结构设计的观点来看,合并和分解都可以在数据项依赖图上直接进行,通常是在单值依赖中,当主键泡中数据项是一切目的泡中数据项的主键时,可以将目的泡合并。如图 5 中学生的“性别”、“学号”、“年级”在图 6 中进行合并表示,这样在重新构成合并表时更为直观。一旦修改了依赖图,则同时也要修改相应依赖表中的依赖语句。

dBASE / FOXBASE 不支持对关键字的描述,数据项依赖图和关系表中标出的主键都只能通过程序中的索引和排序命令的引用才能建立。但这并不说明关键字不重要,相反,主键不仅是建立数据项依赖图的基础,而且主键和外部键是描述数据库结构的数据一致性约束的重要信息。在应用扩充数据项图法建立概念结构模型,的各阶段标明主键,可使用户据此分析有关数据项的操纵频率和依赖关系,从而能恰当进行符合规范化的分解和合并操作。

综上所述,扩充数据项图法简捷直观、易学易懂,特别是提供了一个完成概念结构设计的确思路,在目前微机上一般不具备辅助设计环境,应用 E-R 图法又过于复杂的情况下,是一种值得推荐的实用设计方法。

参考文献:

1. 施伯乐 <<关系数据库的理论及应用>> 河南科技出版社 1989
2. Smith, H. C, Database Design : Compsing Fully Normalized Tables From A Rigorous Dependency Diagram, Communication of the ACM, Vol, 28, pp. 8 (1985)
3. 陈自安 <<应用数据库设计-方法与技术>> 西安电子科技大学出版社 1991
4. 唐常杰 <<实用数据库技术>> 四川大学出版社 1986

