

文章编号:1000-582X(2003)07-0128-04

基于实体连接的信息组织技术*

管昕武¹, 符欲梅¹, 邓琳², 刘飞², 易树平²

(1. 重庆大学 光电技术及系统教育部重点实验室, 重庆 400044; 2. 重庆大学 机械工程学院, 重庆 400044)

摘要:企业在实施信息化过程中,由于未规划系统和信息变动因素等存在,使得企业不得不改变原有信息系统的数据库模式,但现有的信息组织技术及理论却难以解决未规划系统和信息变动对数据库模式无损依赖性和无损连接性的影响,极大地影响了企业实施信息化的进程,也造成了资源浪费。针对以上问题,以关系数据理论为基础,借助E-R语义建模模型,提出了一种基于实体连接的信息组织技术,该信息组织技术具有良好的扩展性和柔性,能够适应企业信息的变化和调整。

关键词:信息组织技术;信息化;关系理论;实体-联系模型
中图分类号:TP391 **文献标识码:**A

以信息技术革命为主导的知识革命促进了传统工业的知识化转变,对企业的生产、经营产生了重大的影响,信息化成为企业发展的必然。企业实施信息化的根本目的是通过提高企业对信息获取、处理、传递、存储和利用的能力来提高企业的竞争能力。信息组织技术着眼于企业信息化全局的信息管理、维护以及存储的构架问题,是企业信息化的基础和根本。企业信息化要求信息组织技术能够适应因企业调整变化而引起的有关信息的调整变化,即考虑未规划系统和信息变动^[1-2]。这里,未规划系统是指目前未纳入考虑或未考虑到的信息;信息变动是指改变原有的信息或信息处理方法。

目前,数据库技术仍然是当前信息组织技术的主流。在数据库基础理论中,关系数据理论仍占有绝对主导地位^[3]。在具体的数据库模式设计方面,有关研究给出了关系模式规范化的分解方法,可以使数据库模式的设计做到近乎完美的地步^[4],但这些方法均建立在既定的“全属性集”的基础之上,当“全属性集”发生改变,即发生信息的调整变化时,原来近乎完美的数据库模式就难以胜任。未规划系统和信息变动对企业信息化工作带来了巨大的负面影响,但目前尚未有有效的解决方案^[5-8]。因此,研究适应未规划系统和信息变动的信息组织技术是当前研究的一个热点方向,具有重要的理论意义和实用价值。

以关系数据理论为基础,借助E-R模型(Entity

- Relationship Model, 实体-联系模型),提出了一种基于实体连接的信息组织技术,能够很好地适应企业信息化中信息的调整与变化。

1 关系数据理论及数据库模式

在关系数据理论中,对数据库模式的定义如下:

对于所关心的客观事物的性质称为属性,属性组成一个有限集合称为一个关系模式。

用 $\mu = \{A, B, C, \dots\}$ 表示在数据库中关心的所有属性的集合,简称“全属性集”,要求 μ 是有限的。其中 A, B, C, \dots 表示客观事物的属性。

若 R_1, R_2, \dots, R_n 都是 μ 的子集,它们都是关系模式,且 $R_1 \cup R_2 \cup \dots \cup R_n = \mu$,则称 $\Omega = \{R_1, R_2, \dots, R_n\}$ 为 μ 上的一个数据库模式。相应, $\omega = \{r_1, r_2, \dots, r_n\}$ 称为 Ω 上的一个关系数据库,其中 r_i 为关系模式 R_i 的实例^[4]。

虽然只要 $R_1 \cup R_2 \cup \dots \cup R_n = \mu, \Omega = \{R_1, R_2, \dots, R_n\}$ 就是 μ 上的一个数据库模式,但存于 μ 上的数据库模式并非唯一,且并非任何一个数据库模式都能满足数据管理的要求。

2 未规划系统和信息变动的分类

在实际应用中,数据模式设计首先借助于语义建模来描述数据的组织。E-R模型是目前一种常用的语

* 收稿日期:2003-03-05

作者简介:管昕武(1969-),男,四川江油人,重庆大学博士,主要从事计算机集成制造系统方向研究。

义建模方法。在E-R模型中,实体是对客观事物或客观事件本身及其属性的描述,联系是对客观事物或客观事件之间联系的描述。

考虑未规划系统和信息变动从本质上看是“全属性集”的变化,是一个处理信息的范围和量的扩展过程。本文从实体和联系的角度将这种扩展分为以下3种情况:

1) 实体属性的扩展

实体属性的扩展是实体本身属性的增加,这种扩展不会影响数据库模式的连接和依赖关系。

2) 实体的扩展

实体的扩展是指在信息管理过程中所关心的客观对象范围的增加,这种扩展会影响原有的数据库模式,可能会破坏数据的完整性和一致性,并影响连接和依赖关系。

3) 联系的扩展

联系的扩展是指在信息管理过程中所关心的客观对象之间的关系的增加,这种扩展一般也会影响原有的数据库模式,也有可能破坏数据的完整性和一致性,并影响原有数据库模式的连接和依赖关系。

从以上3种分类情况可知:实体和联系的扩展会破坏原来的数据库模式,使之产生弊病,因此要求能够适应这两类扩展。

3 基于实体连接的数据模式

从某种意义上讲,E-R建模过程就是一个数据库模式的设计过程。但E-R建模就数据完整性等方面并未予以足够考虑,因此一个采用E-R语义建模的数据库设计直接影射到一个实际的数据库模式上还存在许多问题^[10]。Codd指出,关系是形式化的对象,关系模式是一个形式化的系统,而E-R模型并非一个形式化的模型。本文从关系模式的具体规定开始,研究如何借助E-R语义建模的思想形成一个具体的数据库模式。

3.1 有关实体及其关系模式的规定

实体是本文提出的数据组织技术的基础,规定如下:

实体是关于客观对象是什么或具有什么特征的描述。

1) E_1, E_2, \dots, E_n 分别是描述实体的关系模式, (E_i, Σ_i) 是一组函数依赖模式, $\Sigma_i \vdash X_i \rightarrow Y_i, Y_i \subseteq NPA(E_i, \Sigma_i), X_i$ 是 E_i 唯一的主属性^[4]。($i = 1, 2, \dots, n$)。

2) 且对于任何属性集 $A \subseteq NPA(E_i, \Sigma_i), A$ 必不

包含 $KEY(E_j, \Sigma_j)$, 其中 $i \neq j$ 。($i = 1, 2, \dots, n; j = 1, 2, \dots, n$)。

其中: Σ 表示一个函数依赖集合; $X_i \rightarrow Y_i$ 表示 X_i 函数地决定 Y_i ; $\Sigma_i \vdash X_i \rightarrow Y_i$ 表示 $X_i \rightarrow Y_i$ 被 Σ_i 所蕴涵;

$NPA(E_i, \Sigma_i)$ 表示非主属性集;

$KEY(E_i, \Sigma_i)$ 表示主属性集。

上述规定的含义为:

a. 描述每一实体的关系模式中存在唯一的主属性,且任意非主属性都不传递地依赖于主属性。由于实体是现实的客观存在,都能够有唯一的标识,因此该规定成立。

b. 在任何一个描述实体的关系模式的非主属性中,不存在描述其它实体关系模式的关键字,即描述实体的关系模式中不存在外关键字。该关系模式是属于3NF的关系模式。

3.2 有关联系及其关系模式的规定

联系是对实体之间关系的描述。在现实世界中,实体之间的关系有两种情况:一是对一种状态的描述,如产品和零件的关系就是一种状态;另外是对一个动态事件的描述,如描述顾客购买商品时,还需考虑购买日期等因素,这是一个动态过程。

1) 描述实体之间的一种状态

R 为描述实体 E_1, E_2, \dots, E_n 之间联系的一个关系模式,对于任何属性集 $X \subseteq KEY(R, \Sigma)$, 一定有 $X \subseteq KEY(E_1, \Sigma_1) \cup KEY(E_2, \Sigma_2) \cup \dots \cup KEY(E_n, \Sigma_n)$, 其中 $i = 1, 2, \dots, n$ 。该关系模式属于3NF。

2) 描述实体间的动态事件

R 为描述实体 E_1, E_2, \dots, E_n 之间关系的一个关系模式,其所描述的全部实体的关系模式的主属性只能组成 R 关键字的一部分,即 $KEY(E_1, \Sigma_1) \cup KEY(E_2, \Sigma_2) \cup \dots \cup KEY(E_n, \Sigma_n)$ 是 R 关键字的真子集, R 关键字还需有其它主属性确定。

3.3 基于实体连接的数据模式

根据上面对实体和联系的规定,可给出基于实体连接的数据模式。

1) 描述实体的关系模式

$$E = \{K, A, B, C, \dots, N\}$$

其中 K 是该关系模式的唯一主属性, A, B, C, \dots, N 不是其它任何描述实体的关系模式的主属性。在 E 上的函数依赖集合 Σ 为:

$$\Sigma = \{K \rightarrow A, K \rightarrow B, K \rightarrow C, \dots, K \rightarrow N\}$$

在关系模式 E 中,任何非主属性都不传递地依赖于主属性 K ,因此 E 为第三范式的关系模式,相应 (E, Σ) 属于3NF。

2) 描述联系的关系模式

描述联系的关系模式是描述实体间的的关系。设数据库模式中实体的个数为 n , 每一个关系模式 R 所描述的具有联系实体的个数为 i , 有 $2 \leq i \leq n$ 。

① 描述实体间状态的情况

$$R = \{K, X, Y, \dots, Z\}$$

其中 K 是属性集合 $\{K_1, K_2, \dots, K_n\}$ 的子集。设 K 中的元素个数为 i , 则 $2 \leq i \leq n$ 。 K_1, K_2, \dots, K_n 分别是描述实体的关系模式 E_1, E_2, \dots, E_n 的主属性, K 是 R 的唯一关键字。在 E 上的函数依赖集合 Σ 为:

$$\Sigma = \{K \rightarrow X, K \rightarrow Y, \dots, K \rightarrow Z\}$$

在关系模式 R 中, 任何非主属性都不传递地依赖于 K , 因此 R 为第三范式的关系模式, 相应 (R, Σ) 属于 3 NF。

② 描述实体间动态事件的情况

$$R = \{K, A_1, A_2, \dots, A_j, B_1, B_2, \dots, B_k\}$$

同样, K 是属性集合 $\{K_1, K_2, \dots, K_n\}$ 的子集, 设 K 中的元素个数为 i , 则 $2 \leq i \leq n$ 。关于 K 有如下规定, 对于任何 $\{K_1, K_2, \dots, K_n\}$ 的子集 X, Y , 一定有 $X \neq Y$ 。

K_1, K_2, \dots, K_n 分别是实体 E_1, E_2, \dots, E_n 的主属性, A_1, A_2, \dots, A_j 是关于所描述事件的属性, 如商品销售事件的销售日期、地点等, B_1, B_2, \dots, B_k 是事件的其它属性, K 和 A_1, A_2, \dots, A_j 共同组成关系模式 R 的关键字。在 R 上的函数依赖集合 Σ 为:

$$\Sigma = \{K A_1 A_2 \dots A_j \rightarrow B_1, K A_1 A_2 \dots A_j \rightarrow B_2, \dots, K A_1 A_2 \dots A_j \rightarrow B_k\}$$

在关系模式 R 中, 任何非主属性都不传递地依赖于 $K A_1 A_2 \dots A_j$, 因此 R 为第三范式的关系模式, 相应 (R, Σ) 属于 3 NF。上述关系模式可以进行实体化转化, 变为:

$$E = \{K_0, K, A_1, A_2, \dots, A_j, B_1, B_2, \dots, B_k\}$$

K_0 为事件标识。在 E 上的函数依赖集合 Σ 为:

$$\Sigma = \{K_0 \rightarrow K, K_0 \rightarrow A_1, K_0 \rightarrow A_2, \dots, K_0 \rightarrow A_j, K_0 \rightarrow B_1, K_0 \rightarrow B_2, \dots, K_0 \rightarrow B_k, K A_1 A_2 \dots A_j \rightarrow B_1, K A_1 A_2 \dots A_j \rightarrow B_2, \dots, K A_1 A_2 \dots A_j \rightarrow B_k, K A_1 A_2 \dots A_j \rightarrow K_0\}$$

在关系模式 E 中, 任何非主属性都不传递地依赖于 $K A_1 A_2 \dots A_j$ 和 K_0 ; 所有主属性都完全函数依赖于每个不包含它的关键字, 因此 R 为 BCNF 范式的关系模式, 相应 (R, Σ) 属于 BCNF 范式。

3) 基于实体连接的数据库模式

基于前面给出的实体和联系的一般形式描述, 可以给出基于实体连接的数据库模式一般形式:

对于有限的全属性集 $\mu = \{A, B, C, \dots\}$, 若 $E_1,$

$E_2, \dots, E_n, R_1, R_2, \dots, R_m$ 都是 μ 的子集, 它们都是关系模式, 且 $E_1 \cup E_2 \cup \dots \cup E_n \cup R_1 \cup R_2 \cup \dots \cup R_m = \mu$, 则 μ 上的一个数据库模式 Ω 为:

$$\Omega = \{E_1, E_2, \dots, E_n, R_1, R_2, \dots, R_m\}$$

相应, $\omega = \{r_1, r_2, \dots, r_n\}$ 称为是 Ω 上的一个关系数据库。

4 基于实体连接的数据模式无损连接性、无损依赖性和扩展性分析

根据文中的规定, 每一描述实体和联系的关系模式都是第三范式的关系模式, 相应函数依赖集合属于第三范式; 描述联系实体化转化的关系模式是属于 BCNF 范式的关系模式, 相应函数依赖集合属于 BCNF 范式。因此, 提出的基于实体连接的信息组织技术能满足数据的完整性和一致性约束。

经过理论分析和证明, 提出的基于实体连接的信息组织技术作为一种关系模式分解方法是一种无损连接和无损依赖的分解^[1]。

当“全属性集”发生改变, 即发生信息的调整变化时, 只要全部描述实体关系的关系模式主属性集合不为全部描述实体的关系模式主属性集合的真子集时, 能做到关系模式分解的无损连接性和函数依赖模式分解的无损依赖性^[1]。

5 结论

首先对未规划系统和信息变动进行了分类; 针对未规划系统和信息变动的 3 种情况, 利用关系数据理论和 E-R 建模方法提出了一种信息组织技术, 能够有效保证数据的一致性和完整性等基本要求, 适应企业未来的信息变动; 提出的基于实体连接的信息组织技术为企业信息化提供了一种简单规范的信息组织方法, 对全局数据管理和信息集成以及企业信息化总体规划具有一定指导意义。

提出的基于实体连接的信息组织技术在长安汽车股份有限公司委托项目“长安汽车股份有限公司投资决策支持系统开发研究与项目实施”, 重庆市科学技术委员会科技攻关项目“投资决策支持系统研究”等项目中得到了成功应用, 证明了本文提出的基于实体连接信息组织能够适应未规划系统和信息变动, 满足企业信息化的要求。

参考文献:

- [1] 鲁昕武. 制造企业信息化信息组织相关技术及工程应用 [D]. 重庆: 重庆大学, 2001.

- [2] 商琳, 骆斌. 一种基于数据仓库的数据挖掘系统的结构框架[J]. 计算机应用研究, 2000, 17(9): 63 - 65.
- [3] 陈武, 袁国忠. 企业数据仓库规划建立与实现[M]. 北京: 人民邮电出版社, 2000.
- [4] 吴炜煜. 工程数据库管理系统[M]. 北京: 清华大学出版社, 1996.
- [5] 马垣. 关系数据库理论[M]. 北京: 清华大学出版社, 南宁: 广西科学技术出版社, 1999.
- [6] 张水平, 万映辉. C³I 异构数据库通信平台实现技术[J]. 计算机工程, 2000, 26(2): 100 - 101.
- [7] 齐进. 实现异构数据库集成的一种方法——元数据库法[J]. 计算机工程与应用, 1998, 34(4): 6 - 9.
- [8] 王宁. 异构数据源集成系统查询分解和优化的实现[J]. 软件学报, 2000, 11(2): 222 - 228.
- [9] 朱战立. C/S 方式的单节点分布式数据库设计[J]. 计算机应用研究, 2000, 17(6): 30 - 31.
- [10] C J DATE. 数据库系统导论[M]. 孟小峰, 王珊译. 北京: 机械工业出版社, 2000.

Information Organization Technique Based on Entity Linkage

ZAN Xin-wu¹, FU Yu-mei¹, DENG Lin², LIU Fei², YI Shu-ping²

(1. Key Laboratory of Optoelectronic Technology & Systems under the State

Ministry of Education, Chongqing University, Chongqing 400044, China;

2. College of Mechanical Engineering, Chongqing University, Chongqing 400044, China)

Abstract: In the course of Information Engineering of enterprises, because of the existence of unplanned system and information verification, enterprises have to change primary data model of the information system. But existing information organizing technique and theory can not solve the influence that unplanned system and information verification impact on preserve linkage and preserve dependency of the data model, and this has great effect on the process of Information Engineering. Aiming at solving the problem mentioned above, on the basis of Relation Data Theory and making use of E - R model, the paper puts forward an information organization technique based on entity linkage with well expandability and flexibility. The technique can adapt to information verification and adjustment of enterprise.

Key words: information technology; information engineering; relation data theory; entity - relationship model

(编辑 张小强)