

# 数据库逻辑设计的 E-S-C 转换方法\*

曾 平

(武汉测绘科技大学, 计算机科学与工程系, 武汉市珞喻路39号, 430070)

**摘 要** 本文提出了一个基于实体-联系模型(ERM)的数据库逻辑设计方法, 即由 EERM 到 SERM, SERM 到 CERM 的转换方法, 简称 E-S-C 方法。首先介绍了一个扩充实体-联系模型(EERM), 定义了简单实体-联系模型(SERM)及规范实体-联系模型(CERM), 然后给出了 EERM 到 SERM, SERM 到 CERM 的等价转换规则。

**关键词** 数据库, 计算机辅助设计, 扩充实体-联系模型

**分类号** TP311.13, TP302.2

数据库设计分为需求分析、逻辑设计和物理设计三个阶段<sup>[1,2]</sup>。逻辑设计是数据库设计的中心环节, 很大程度上决定着一个数据库系统的好坏。它存在两个主要问题: ①逻辑设计是一个劳动强度大、费时多的过程; ②设计的质量与设计者的经验和素质有关, 无法保证最优。为此本文提出 E-S-C 转换方法, 以期解决上述两个问题。

实体-联系模型(Entity-Relationship Model, 简称 ERM)于1976年由 P. P. Chen 首次提出<sup>[3]</sup>。它克服了网络、层次及关系模型的不足, 并继承了它们的大部分优点。因 ERM 表现手法简单, 易于理解, 又具有高度的数据独立性, 使其在数据库设计方面得到了广泛的应用。由于 ERM 对现实世界的描述手段不够充分, 特别是在语义描述方面, 因而需对其扩充。下文中的 EERM (Extended ERM)正是为此目的而设计的<sup>[4]</sup>。EERM 主要着眼于为数据库设计者提供更加丰富、自然的描述手段, 而没有考虑所表达数据的冗余及规范化程度。为此又定义了一个 SERM (Simple ERM), 它的规则数据库模式是1NF的。然后, 又给出了 CERM(Canonical ERM)的定义。CERM 具有较高的规范化程度, 从 CERM 向经典的数据库模型转换既方便又可靠。因此, 实现从 EERM 到 CERM 的转换就可获得规范化程度较高的企业模式, 一方面完成了数据库逻辑设计, 另一方面又能较好地解决数据库逻辑设计中存在的问题。

## 1 扩充实体-联系模型的定义

EERM 是对 ERM 的扩充, 其组成要素仍为实体、联系和属性。下面给出其定义及分类。

### 1.1 实体的定义及分类

定义: 具有相同性质的一类实体的集合称为实体集, 实体集中的一个元素称为实体或实体实例。形如  $E(A_1, A_2, \dots, A_n)$  的结构称为实体型。其中,  $E$  为实体集名,  $A_1, A_2, \dots, A_n$  为  $E$  的属性集。

在下文中, 只要不引起混淆, 有时用实体代替实体集或实体型。

定义: 如果实体集  $A$  中的每个实体的存在依赖于实体集  $B$  中一个或多个实体的存在, 则

收稿日期: 1993-09-24。曾平, 女, 29岁, 讲师, 现从事地图汉字注记专家系统的研究。

\* 国家测绘局“八五”科技攻关资助项目。

称实体集  $A$  存在依赖于实体集  $B$ 。称  $A$  为依赖实体,  $B$  为独立实体。

定义:如果实体集  $E_1$  中的每个实体,也同时是实体集  $E_2$  中的一个实体,则称  $E_1$  是  $E_2$  的子集,  $E_2$  为  $E_1$  的超集,实体集  $E_1$ 、 $E_2$  之间存在子集层次。

定义:设  $E$  为一实体集,  $E_1, \dots, E_n$  均为  $E$  的子集,且  $E_i \cap E_j = \varnothing, 1 \leq i \neq j \leq n, \bigcup_{i=1}^n E_i = E$ , 则称  $E$  是实体集合  $E_1, \dots, E_n$  的归纳。 $E$  称为超集,实体集  $E$  与实体集合  $E_1, E_2, \dots, E_n$  之间存在归纳层次。

显然,  $E_1, E_2, \dots, E_n$  构成了  $E$  的一个聚类,这种聚类划分通常是根据  $E$  的一个属性来进行的。此属性称为  $E$  的判别属性。

独立实体和依赖实体分别

用单边矩形和双边矩形表示,子集用带箭头的矩形表示,它们的图形表示如图1所示。

### 1.2 联系的定义及分类

定义:联系集  $R$  是几个实体集之间的一种语义关系。 $R$  中的一个元素称为联系或联系实例,形如  $R(E_1, \dots, E_n, A_1, \dots, A_m)$  的结构称为联系型。其中,  $R$  是联系集名,  $A_1, \dots, A_m$  为  $R$  自身的属性,  $m \geq 0, n \geq 1$ 。

在不致混淆的情况下,以后的叙述中用联系代替联系集或联系型。

与实体分类相对应,联系也分为两类。

定义:设  $R$  是定义在  $n+1$  个实体  $E_1, \dots, E_n, E$  上的联系,如果  $E$  通过  $R$  存在依赖于实体集  $E_1, \dots, E_n$ , 则称  $R$  为依赖联系,否则称  $R$  为独立联系。

定义:一个联系所关联实体的数目称为联系的度。度为  $n$  的联系称为  $n$  元联系。

定义:联系  $R$  定义在实体集  $E_1, \dots, E_n$  上,如给定除  $E_i$  以外的所有实体集中的某个实体,最多只有  $E_i$  中的一个实体与之相联系,则称  $E_i$  相对于  $R$  的映射值为1,否则称  $E_i$  相对于  $R$  的映射值为多。

定义:联系集  $R$  定义在实体集  $E_1, \dots, E_n$  上。如  $E_i$  中的每个实体必须在  $R$  的某个联系中出现,则称  $E_i$  相对于  $R$  的成员籍为必选,否则称  $E_i$  相对于  $R$  的成员籍为可选。

独立联系和依赖联系分别用单菱形框和双菱形框表示。实体相对于联系的映射及成员籍直接标明在联系与实体之间的连线上。其图形表示法如图1所示。

### 1.3 属性及实体标识的定义

定义:属性是事物某方面的特征,可以形式地定义为从实体或联系到值集的映射。

实体或联系到值集的映射有以下两种情况:

1( $n$ ):1:实体或联系集中的一个实例仅与值集中的一个值对应,而值集中的一个值与实体或联系集中的一个(多个)实例对应。

1( $n$ ): $n$ :实体或联系集中的一个实例与值集中的多个值对应,而值集中的一个值与实体或

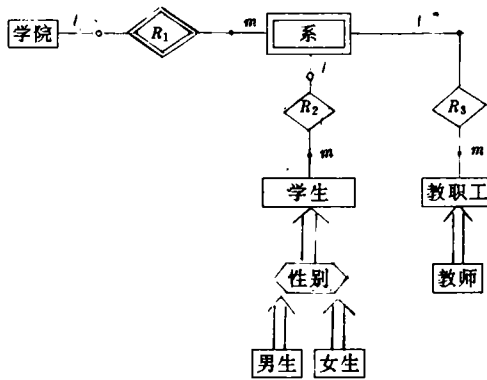


图1 实体及联系的图形表示法

联系集中的一个(多个)实例对应。

定义:如果一个属性将一个实体或联系映射到值集中的一个值,则称该属性为单值属性,否则称该属性为多值属性。

定义:聚集是一个属性集合,此集合可以看作单一特性。

定义:一个实体集中用以区别各个实体的一个或多个属性称为实体标识。通常选择一个有意义而且简单的标识作为实体的主标识,其余的标识称为候选标识。出现在实体标识中的属性称为主属性,其余的属性称为非主属性。如果实体标识中的所有属性均为该实体本身的属性,则称该实体的标识为内部标识,否则称为外部标识。

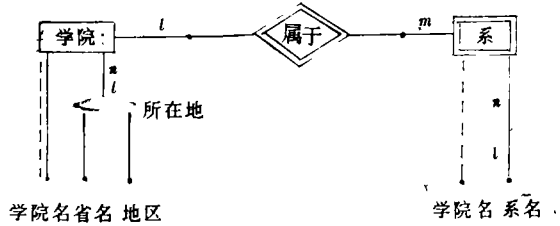


图2 属性及实体标识的图形表示法

属性及实体标识的图形表示法如图2所示。

## 2 SERM 及 CERM 的定义

前述的 EERM 主要着眼于为数据库设计者提供更丰富的语义描述手段,而未考虑所表达数据的冗余、实体及联系的规范化程度等。为解决上述问题,引入 SERM 及 CERM 的定义。

定义:设  $\epsilon$  是一个 ERD(Entity-Relationship Diagram),  $\epsilon$  的规则数据库模式  $N$  由下列关系模式组成:

- i) 对  $\epsilon$  中的每个实体  $E$ , 对应的关系模式包含  $E$  的所有属性。
- ii) 对  $\epsilon$  中的每个联系  $R$ , 它定义在  $E_1, \dots, E_k$  上, 对应的关系模式由  $R$  自身的属性加上  $E_1, \dots, E_k$  的主标识组成。而  $\epsilon$  的泛关系模式  $u$  由  $N$  中所有关系模式的属性集合的并组成。

SERM 是 EERM 的一个子集。与 EERM 相比,它对实体、联系及属性有了更多的限制:

- 实体只划分为独立实体及依赖实体,其定义及图形表示法同 EERM。
- 联系均为二元联系,且划分为独立联系和依赖联系。联系的度、实体相对于联系的映射、成员籍、独立联系、依赖联系的定义及图形表示法同 EERM。
- 属性只能为单值属性,并区分为主属性和非主属性,并且不存在聚集,其定义及图形表示法同 EERM。
- 实体标识的定义及图形表示法同 EERM。

定理:设  $\epsilon$  是一个 SERD(Simple ERD), 则  $\epsilon$  的规则数据库模式  $N$  是 1NF 的。

定义:如果一个实体所对应的关系模式是 3NF 的,则称它为正规实体;如果一个联系所对应的关系模式是 3NF 的,则称它为正规联系。

CERM 由正规实体、正规联系及单值属性组成,其定义如下:

- 实体分为独立实体及依赖实体,但二者均为正规实体,其定义及图形表示法同 EERM。
- 联系分为二元联系和三元联系,但二者均为正规联系。联系的度、实体相对于联系的映射、成员籍、独立联系、依赖联系的定义及图形表示法同 EERM。
- 属性只能为单值属性,并区分为主属性和非主属性,其定义及图形表示法同 EERM。
- 实体标识的定义及图形表示法同 EERM。

定理:设  $\epsilon$  是一个 CERD(Canonical ERD), 则  $\epsilon$  的规则数据库模式  $N$  是 3NF 的。

### 3 ERD 等价概念

ERM 的优点之一是其定义的清晰性及语义的可见性。在一个 ERD 中除了模型定义中的实体、联系及属性外,还蕴涵有嵌入 ERD 的函数依赖<sup>[5]</sup>。

自然嵌入在 ERD 上的函数依赖有以下几种(对 SERD 及 CERD 而言):

a) 给定一个实体关系模式  $E(K_1, \dots, K_m, A_1, \dots, A_n)$ , 其中  $K_1$  是实体的主标识,  $K_2, \dots, K_m$  是候选标识,  $A_i$  为单值非主属性,  $m > 0, n \geq 0, 1 \leq i \leq n$ 。则自然嵌入到  $E$  中的函数依赖有:

$$K_j \rightarrow K_l \quad 1 \leq i \neq l \leq m, \quad K_j \rightarrow A_i \quad 1 \leq i \leq n$$

b) 给定一个联系关系模式  $R(K_1, K_2, D_1, \dots, D_u)$ , 其中  $K_1, K_2$  是与  $R$  相关实体  $E_1, E_2$  的主标识,  $D_1, \dots, D_u$  是联系本身的属性,  $u \geq 0$ 。

i) 如  $E_1, E_2$  相对于  $R$  的映射值均为 1, 则自然嵌入到  $R$  中的函数依赖有:

$$K_1 \rightarrow K_2 \quad K_2 \rightarrow K_1 \quad K_i \rightarrow D_j \quad 1 \leq i \leq 2, \quad 1 \leq j \leq u$$

ii) 如  $E_1, E_2$  中只有一个实体相对于  $R$  的映射值为 1, 不妨设  $E_2$  相对于  $R$  的映射值为 1,  $E_1$  相对于  $R$  的映射值为多, 则自然嵌入到  $R$  中的函数依赖有:

$$K_1 \rightarrow K_2 \quad K_1 \rightarrow D_i \quad 1 \leq i \leq u$$

iii) 如  $E_1, E_2$  相对于  $R$  的映射值均为多, 则自然嵌入到  $R$  中的函数依赖有:

$$K_1 K_2 \rightarrow D_i \quad 1 \leq i \leq u$$

定义: 令  $e$  与  $e'$  是 2 个 ERD, 对应的规则数据库模式为  $N$  和  $N'$ , 泛关系模式为  $u$  及  $u'$ ,  $F, F'$  是  $u, u'$  上的函数依赖集(由自然嵌入在实体、联系上的函数依赖集和非自然嵌入在实体、联系上的函数依赖集的并组成)。如果  $u = u', F \equiv F'$  且  $\{I | I \text{ 是 } u \text{ 的一个实例, } I \text{ 满足 } F \text{ 且 } I = m_u(I)\} = \{I' | I' \text{ 是 } u' \text{ 的一个实例, } I' \text{ 满足 } F' \text{ 且 } I' = m_{u'}(I')\}$ , 则称  $e$  与  $e'$  实例数据等价。

### 4 EERD 到 CERD 的转换

为消除 EERD 中的数据冗余,使其达到较高的规范化程度,需要实现 EERD 到 CERD 的等价转换。限于篇幅,转换前后 ERD 的等价性证明略。

#### 4.1 EERD 到 SERD 的转换

(1) 消除子集与归纳层次: 设子集集合为  $S, S_i \in S, 1 \leq i \leq n$ , 超集为  $E$ , 则

- 对每个  $S_i \in S$ , 建立一个依赖实体  $E_i$ ,  $E_i$  的属性及映射类型同  $S_i$ ,  $E_i$  的标识为  $E$  的标识, 标明在  $E_i$  上的函数依赖集同标明在  $S_i$  上的函数依赖集(即非自然嵌入  $S_i$  上的函数依赖集, 以下简称函数依赖集)。

- 在每个  $E_i$  与  $E$  之间建立一个二元依赖联系  $R_i$ ,  $E$  相对于  $R_i$  的映射为 1, 成员籍为可选;  $E_i$  相对于  $R_i$  的映射为 1, 成员籍为必选。  $R_i$  无属性。

- 将每个  $S_i \in S$  从  $E$  中删去。

(2) 消除聚集: 设在实体或联系中, 聚集的集合为  $AG, AG_i \in AG, 1 \leq i \leq n, AG_i$  的分量属性集合为  $\{A_{ij}\}, 1 \leq i \leq m_i$ , 则

- 对每个  $AG_i \in AG$ , 将  $AG_i$  的各分量属性直接加入到它所属的实体及联系中。  $A_{ij}$  的映射类型为  $n:1$ , 属性类别同  $AG_i, 1 \leq i \leq n, 1 \leq j \leq m_i$ 。

- 将所有  $AG_i \in AG$  从它所属的实体或联系中删去, 并将实体或联系上的函数依赖集中的

AG, 用  $\{A_i\}$  替代。

(3) 消除多值属性: 设实体  $E$  中,  $mv$  为多值属性,  $Fu$  为  $E$  上的函数依赖集,  $K$  为  $E$  的主标识, 则

- 建立一个独立实体  $E_1$ , 仅有单个属性  $mv$ .  $mv$  的映射类型为  $1:1$ ,  $E_1$  的标识为  $mv$ .
- 在  $E$  与  $E_1$  之间建立一个独立二元联系  $R_1$ .  $E$  相对于  $R_1$  的映射为  $1$ , 成员籍为必选;  $E_1$  相对于  $R_1$  的映射为  $m$ , 成员籍为必选.
- 将  $mv$  从  $E$  的属性集中删去, 将  $Fu$  中包含  $mv$  的函数依赖从  $Fu$  中删去.

(4) 多元联系到二元联系的转换: 设  $n$  元联系  $R$  定义在  $E_1, \dots, E_n$  上,  $K_1, \dots, K_n$  分别为  $E_1, \dots, E_n$  的主标识, 则

• 生成联系实体  $ER$ , 其属性由  $K_1, \dots, K_n$  与  $R$  自身的属性组成, 标识为  $K_1, \dots, K_n$ . 当  $E_1, \dots, E_n$  相对于  $R$  的映射值均为  $m$  时,  $ER$  上的函数依赖集同  $R$  上的函数依赖集; 当  $E_1, \dots, E_m$  相对于  $R$  的映射值为  $1, E_{m+1}, \dots, E_n$  相对于  $R$  的映射值为  $m$  时,  $ER$  上的函数依赖集为  $R$  上的函数依赖集与  $\{K_1 \dots K_{i-1} K_{i+1} \dots K_m K_{m+1} \dots K_n \rightarrow K_i \mid 1 \leq i \leq m, m < n\}$  的并。

• 在每个  $E_i$  与  $ER$  之间建立一个二元独立联系  $R_i, 1 \leq i \leq n$ .  $ER$  相对于  $R_i$  的映射为  $m$ , 成员籍为必选;  $E_i$  相对于  $R_i$  的映射为  $1$ , 成员籍同  $E_i$  相对于  $R$  的成员籍。

• 删去  $R$ .

### 4.2 SERD 到 CERD 的转换

(1) 预处理: 求出 SERD 中各对象上的最小函数依赖集, 确定联系实体上的标识, 按函数依赖集修改联系的映射类型。

(2) 将实体转换为正规实体: 设实体关系为  $E(x_1, \dots, x_n, Y)$ , 其中  $x_1$  为主标识,  $x_2, \dots, x_n$  为候选标识,  $Y$  为非主属性集,  $Fu$  为  $E$  上的函数依赖集, 则

i) 若  $Fu$  为空, 则  $E$  已为正规实体。

ii) 若  $Fu$  不为空, 则对  $Fu$  按具有等价左部的原则进行分组。假定  $Fu$  分为  $K$  组, 对每一组函数依赖  $F_i = \{A_{i1} \rightarrow B_{i1}, \dots, A_{im_i} \rightarrow B_{im_i}\}, u_i = \bigcup_{j=1}^{m_i} A_{ij} \cup \bigcup_{j=1}^{m_i} B_{ij}$

• 建立一个独立实体  $E'_i(u_i)$ , 其标识为  $A_{i1}, \dots, A_{im_i}, A_{i1}$  为主标识。如  $A_{ij}$  为单个属性, 则  $A_{ij}$  的映射类型为  $1:1$ , 其余属性的映射类型均为  $n:1$ 。

• 在  $E$  与  $E'_i$  之间引入一个独立二元联系  $R'_i$ .  $E$  相对于  $R'_i$  的映射为  $m$ , 成员籍为必选;  $E'_i$  相对于  $R'_i$  的映射为  $1$ , 成员籍为必选。

• 将所有  $\bigcup_{i=1}^K u_i \cap Y$  中的属性从  $E$  中删去, 并置  $Fu$  为空。

(3) 将联系转换为正规关系: 设联系  $R$  定义在  $E_1, E_2$  上,  $K_1, K_2$  是  $E_1, E_2$  的主标识, 联系关系为  $R(K_1, K_2, Y)$ . 其中  $Y$  是  $R$  自身的属性,  $Fu$  是  $R$  上的函数依赖集, 则

i) 若  $Fu$  为空, 则  $R$  已为正规联系。

ii) 若  $Fu$  中存在形如  $K_j \rightarrow Y'_j (j=1, 2)$ , 则

- 将  $Y'_j$  中的属性加入到实体  $E_j$  中作为非主属性,  $Y'_j$  中属性的映射类型为  $n:1$ 。
- 将  $K_j \rightarrow Y'_j$  从  $Fu$  中删去。

iii) 若  $Fu$  不为空, 则对  $Fu$  按具有等价左部的原则进行分组。假定分为  $K$  组, 对每一组函数依赖  $F_i = \{A_{i1} \rightarrow B_{i1}, \dots, A_{im_i} \rightarrow B_{im_i}\}, u_i = \bigcup_{j=1}^{m_i} A_{ij} \cup \bigcup_{j=1}^{m_i} B_{ij}$ , 则

• 建立一个独立实体  $E'_i(u_i)$ , 实体标识为  $A_{i1}, \dots, A_{im_i}, A_{i1}$  为主标识。如果  $A_{ij}$  为单个元素, 则  $A_{ij}$  的映射类型为  $1:1$ , 其余属性的映射类型均为  $n:1$ 。

·如果  $E_1, E_2$  相对于  $R$  的映射均为  $m$ , 则在  $E_1, E_2, E_1'$  之间引入一个三元独立联系。 $E_1'$  相对于  $R_1'$  的映射为 1, 成员籍为必选;  $E_1$  相对于  $R_1'$  的映射为  $M$ , 成员籍同  $E_1$  相对于  $R$  的成员籍;  $E_2$  相对于  $R_1'$  的映射为  $M$ , 成员籍同  $E_2$  相对于  $R$  的成员籍。

·如果  $E_1, E_2$  中至少有一个实体相对于  $R$  的映射为 1, 不妨设该实体为  $E_2$ , 则在  $E_1$  与  $E_1'$  之间引入一个独立二元联系  $R_1'$ 。 $E_1$  相对于  $R_1'$  的映射为  $M$ , 成员籍为  $E_1$  相对于  $R$  的成员籍;  $E_1'$  相对于  $R_1'$  的映射为 1, 成员籍为必选。

iv) 将  $(\bigcup_{i=1}^2 Y_i' \cup \bigcup_{i=1}^k U_i) \cap Y$  中的属性从  $R$  中删去, 置  $F_u$  为空。

本文所述的 E-S-C 方法主要考虑的是 ERD 上静态数据的等价变换, 因此在课题的进一步研究中还需对 ERD 上动态行为的描述及等价变换等问题进行探索。

### 参 考 文 献

- 1 萨师煊, 王珊. 数据库系统概论. 北京: 高等教育出版社, 1983
- 2 霍列基威茨. 数据库分析与设计. 周佩德译. 南京: 南京工学院出版社, 1987
- 3 Chen P P. The Entity-Relationship Model: Toward a Unified View of Data. ACM Transactions on Database Systems, 1976, 1(1): 9~36
- 4 Victor M, Shoshai A. Representing Extended Entity-Relationship Structures in Relational Database: A Modular Approach. ACM Transactions on Database Systems, 1991, 16(3): 423~465
- 5 Jajodia S. The Program of Equivalence for Entity-Relationship Diagrams. IEEE Transactions on Software Engineering, 1983, 9(5): 37~50

## The E-S-C Transformation Method of Logic Database Design

Zeng Ping

(Dept. of Computer Science and Engineering, WTUSM, Luoyu Road 39, Wuhan, China, 430070)

**Abstract** In this paper, a method of logic database design—E-S-C transformation method is presented, which is based on entity-relationship model. Each step of the method can be implemented by algorithm, which is convenient for developing logic database system. An extended entity-relationship model is described, and a simple entity-relationship model and a canonical entity-relationship model are defined. The transformation rules from extended entity-relationship model to simple entity-relationship model is given, then the transformation rules from simple entity-relationship model to canonical entity-relationship model. An automatic tool of logic database design using E-S-C method has been developed.

**Key words** database; computer-aid design; extended entity-relationship model