

国外计算机科学教材系列

信息建模与 关系数据库

—— 从概念分析到逻辑设计

Information Modeling and Relational Databases:
from Conceptual Analysis to Logical Design

TERRY HALPIN

Information
Modeling and
Relational
Databases

FROM CONCEPTUAL ANALYSIS TO LOGICAL DESIGN

Using ORM with ER and UML

[美] Terry Halpin 著

施伯乐 刘国华 陈子军 等译

M K MORGAN
KAUFMANN



电子工业出版社
Publishing House of Electronics Industry
<http://www.phei.com.cn>

国外计算机科学教材系列

信息建模与关系数据库

——从概念分析到逻辑设计

Information Modeling and Relational Databases:
from Conceptual Analysis to Logical Design

[美] Terry Halpin 著

施伯乐 刘国华 陈子军 等译

电子工业出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

本书主要讲解了对象角色建模及其相关的技术知识。书中从ORM角度介绍了实体联系建模(ER)和统一建模语言(UML)。本书还清晰地涵盖了关系数据库理论和SQL的最新进展情况,并阐述了所需的数学和集合理论的概念。事实上,它是此领域惟一一本超出介绍性层面的书,涵盖了把领域专家的知识转化为合理的数据库设计的深入知识。本书通过实例、练习和背景知识介绍,将一步步地教会你如何开发一个基于自然语言的ORM模型,然后,在需要的地方从中提取ER和UML模型。通过学习本书,你会很快精通建模技术,准确高效地开发符合商业目标的数据库。

本书可作为大专院校计算机相关专业高年级本科生及研究生的教材,也适合于从事数据库相关研究的信息建模人员、系统分析员、数据库设计及管理人員等。

Authorized translation from the English language edition published by Elsevier Science(USA). Copyright © 2001 by Academic Press.

Translation Copyright © 2004 by Publishing House of Electronics Industry. All rights reserved.

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from the Publisher.

本书中文简体专有翻译出版版权由Elsevier Science(USA)授予电子工业出版社。其原文版权及中文翻译出版版权受法律保护。未经许可,不得以任何形式或手段复制或抄袭本书内容。

版权贸易合同登记号 图字:01-2002-3815

图书在版编目(CIP)数据

信息建模与关系数据库——从概念分析到逻辑设计/(美)哈尔平(Halpin, T.)著;施伯乐等译.-北京:电子工业出版社,2004.7

(国外计算机科学教材系列)

书名原文:Information Modeling and Relational Databases: from Conceptual Analysis to Logical Design

ISBN 7-121-00074-1

I. 信... II. ①哈... ②施... III. 关系数据库-系统建模-教材 IV. TP311.138

中国版本图书馆CIP数据核字(2004)第063744号

责任编辑:陶淑毅

印 刷:北京兴华印刷厂

出版发行:电子工业出版社

北京市海淀区万寿路173信箱 邮编:100036

经 销:各地新华书店

开 本:787×1092 1/16 印张:34.5 字数:971千字

印 次:2004年7月第1次印刷

定 价:55.00元

凡购买电子工业出版社的图书,如有缺损问题,请向购买书店调换;若书店售缺,请与本社发行部联系。联系电话:(010)68279077。质量投诉请发邮件至zlts@phei.com.cn,盗版侵权举报请发邮件至dbqq@phei.com.cn。

出版说明

21 世纪初的 5 至 10 年是我国国民经济和社会发展的关键时期,也是信息产业快速发展的关键时期。在我国加入 WTO 后的今天,培养一支适应国际化竞争的一流 IT 人才队伍是我国高等教育的重要任务之一。信息科学和技术方面人才的优劣与多寡,是我国面对国际竞争时成败的关键因素。

当前,正值我国高等教育特别是信息科学领域的教育调整、变革的重大时期,为使我国教育体制与国际化接轨,有条件的高等院校正在为某些信息学科和技术课程使用国外优秀教材和优秀原版教材,以使我国在计算机教学上尽快赶上国际先进水平。

电子工业出版社秉承多年来引进国外优秀图书的经验,翻译出版了“国外计算机科学教材系列”丛书,这套教材覆盖学科范围广、领域宽、层次多,既有本科专业课程教材,也有研究生课程教材,以适应不同院系、不同专业、不同层次的师生对教材的需求,广大师生可自由选择和自由组合使用。这些教材涉及的学科方向包括网络与通信、操作系统、计算机组织与结构、算法与数据结构、数据库与信息处理、编程语言、图形图像与多媒体、软件工程等。同时,我们也适当引进了一些优秀英文原版教材,本着翻译版本和英文原版并重的原则,对重点图书既提供英文原版又提供相应的翻译版本。

在图书选题上,我们大都选择国外著名出版公司出版的高校教材,如 Pearson Education 培生教育出版集团、麦格劳-希尔教育出版集团、麻省理工学院出版社、剑桥大学出版社等。撰写教材的许多作者都是蜚声世界的教授、学者,如道格拉斯·科默(Douglas E. Comer)、威廉·斯托林斯(William Stallings)、哈维·戴特尔(Harvey M. Deitel)、尤利斯·布莱克(Uyless Black)等。

为确保教材的选题质量和翻译质量,我们约请了清华大学、北京大学、北京航空航天大学、复旦大学、上海交通大学、南京大学、浙江大学、哈尔滨工业大学、华中科技大学、西安交通大学、国防科学技术大学、解放军理工大学等著名高校的教授和骨干教师参与了本系列教材的选题、翻译和审校工作。他们中既有讲授同类教材的骨干教师、博士,也有积累了几十年教学经验的老教授和博士生导师。

在该系列教材的选题、翻译和编辑加工过程中,为提高教材质量,我们做了大量细致的工作,包括对所选教材进行全面论证;选择编辑时力求达到专业对口;对排版、印制质量进行严格把关。对于英文教材中出现的错误,我们通过与作者联络和网上下载勘误表等方式,逐一进行了修订。

此外,我们还将与国外著名出版公司合作,提供一些教材的教学支持资料,希望能为授课老师提供帮助。今后,我们将继续加强与各高校教师的密切联系,为广大师生引进更多的国外优秀教材和参考书,为我国计算机科学教学体系与国际教学体系的接轨做出努力。

电子工业出版社

教材出版委员会

- | | | |
|-----|-----|---|
| 主 任 | 杨芙清 | 北京大学教授
中国科学院院士
北京大学信息与工程学部主任
北京大学软件工程研究所所长 |
| 委 员 | 王 珊 | 中国人民大学信息学院院长、教授 |
| | 胡道元 | 清华大学计算机科学与技术系教授
国际信息处理联合会通信系统中国代表 |
| | 钟玉琢 | 清华大学计算机科学与技术系教授
中国计算机学会多媒体专业委员会主任 |
| | 谢希仁 | 中国人民解放军理工大学教授
全军网络技术研究中心主任、博士生导师 |
| | 尤晋元 | 上海交通大学计算机科学与工程系教授
上海分布计算技术中心主任 |
| | 施伯乐 | 上海国际数据库研究中心主任、复旦大学教授
中国计算机学会常务理事、上海市计算机学会理事长 |
| | 邹 鹏 | 国防科学技术大学计算机学院教授、博士生导师
教育部计算机基础课程教学指导委员会副主任委员 |
| | 张昆藏 | 青岛大学信息工程学院教授 |

译者序

信息建模是构建信息系统最重要的一步,它可以帮助人们创建出清晰准确、能够代表用户论域(UoD)的有效形式模型。信息系统离不开数据库的支持,数据库是信息系统的重要基础设施。为了保证数据库中的数据正确、完整并被有效地访问,需要对数据库进行合理的设计。信息建模所提供的有关论域的有效形式模型,为合理的数据库设计提供了保障。目前,主要的信息建模方法有:实体-联系建模(ER)、面向对象建模和面向角色建模(ORM)。

实体-联系建模(ER)是由 Peter Chen 于 1976 年提出的,是至今仍被广泛使用的信息建模方法。它依照具有属性和联系的实体来描述世界。长期以来,出现了许多不同版本的 ER,但是至今,仍没有标准的 ER 符号。不同版本的 ER 支持不同的概念并且可能相同的概念使用不同的符号。

面向对象建模是一种将数据和行为都封装到对象里的方法。尽管它主要用来为面向对象编程设计代码,但也能用于数据库设计。目前存在许多面向对象方法,但是至今最有影响力的是被对象管理组(OMG)接受的统一建模语言(UML)。UML 用类图来表示静态数据结构、操作以及对特定的面向对象代码(例如,属性的可视性和关联的导航性)的低级设计决策。UML 的类图可以看做是 ER 图的扩展版本。

本书所介绍的面向角色建模(ORM)依照自然概念(如对象和角色)来表示模型。由于使用自然语言、直观图、实例和依照单一或基本的事实来检验信息,从而简化了分析和设计步骤,提供了一个真正的概念上的建模方法。

与 ORM 模型相比,ER 模型和 UML 模型均不适于形式表示、转换或设计一个概念信息模型的任务。ER 图和 UML 图与自然语言相差甚远,缺乏用于约束的基于角色符号的表达能力和简单性。在域的演化方面不稳定,难以和事实实例并存,并且常常隐藏附有模型的语义域的信息。对于概念信息分析,ORM 比 ER 图及 UML 方法有一些优势。例如,ORM 模型易于语言表示和成员化以便于领域专家的确认,应用域变化时,它们比较稳定,而且它们能够在图表中得到更多的业务规则。然而,ER 图和 UML 类图适于简明的总结。它们的结构更接近于最终的数据库实现,所以,也是有价值的。因此,本书在重点介绍面向角色建模(ORM)的基础上,又涵盖了用 ER 和 UML 进行信息建模的章节,并说明了如何能够从 ORM 模型中容易地提取出 ER 和 UML 数据模型。

全书深入阐述了把领域专家的知识转化为合理的数据库设计所需要的方法。除此之外,还提供了一个简单的概念框架以理解数据库系统到底是什么,并对 SQL 进行了详尽的介绍。它不仅适用于数据库专业人员、计算机科学和信息管理人员,也适用于那些想以人们容易理解且在计算机上容易实现的方式表述应用信息结构的人员。

施伯乐教授对本书的翻译工作进行了细心的指导。全书的翻译工作主要由刘国华、陈子军和王颖完成。胡小明、马朝虹、李晶、宋驰、刘佳、张颖、方为、车建华、艾鸿、马博、程蕾等同学参与了初期的翻译工作。施伯乐教授和刘国华教授负责全书的统稿、审校和组织工作。

在翻译过程中,由于译者水平所限,书中难免有不妥之处,敬请读者批评指正。

序一

John Zachman

Zachman 国际集团创始人，总裁

我与 Terry Halpin 相识多年，早在结识他本人之前就已经知道 Terry Halpin 的大名。虽然他的名气很大，但我认为他是那种既有真才实学又非常谦虚的学者。

当 Terry 邀我为他的新书作序时，我的第一反应就是：“哎呀，我有资格为 Terry Halpin 的书作序吗？”在收到他送来的手稿后，我接下来的反应就是自己竟然难以放下这本书了！你能想到吗？一本技术性的书籍竟会让人如此爱不释手！

的确，这是一本技术性的书籍，Terry 在书中详细探讨了怎样通过建立图解模型来准确而严谨地抓住一个信息系统中语义的细微差异。同时这也是一本通俗易懂的书籍，行文简洁流畅，解释深入浅出。Terry 通过示例和测试来增进你的理解。根据示例建立的众多模型使你无需繁复的过程即可直达问题的要害和核心。

在这本书中，Terry 使用对象-角色建模（ORM）来阐述示例，不仅因为他事实上是基于现代 ORM 来写书，更是因为 ORM 无可比拟的捕获语义意图和以图形清晰表达的能力，为了使建模者掌握 ORM 语言的性能，Terry 极为详尽地探讨了 ORM 建模。然而，实例中也有很多分析，从而使得即使是对 ORM 不熟悉的建模者也能准确地抽取出“论域”的精确语义。

但是我认为，上述所有这些特征都不能真正说明本书的精髓所在。这本书的精髓体现在两个方面。首先，它清晰而鲜明地证明了，准确识别和获取一个企业的意图并将其实现是一系列难以置信的复杂过程。毋庸置疑，在过去的 50 年中（也就是“数据处理”的全部历史）我们所实现的都是缺乏灵活性、适应性差、方向偏移和分裂的系统，它不但反应迟钝而且价格昂贵、难以维护，对管理造成了干扰。而我们总是要不厌其烦地产生准确的概念模型！如果你不严格地描述一个企业开始，那么，又怎么能去指望其他人提出能反映出企业的现实或潜在意图，且随着时间推移而不断适应新变化的设计和实现呢？

遗憾的是，很少有经理人愿意考虑正式地来读一下像《信息建模与关系数据库——从概念分析到逻辑设计》这一类的技术性书籍。但是所有的经理人的确应该来读读这种书，以对他的企业在语义复杂性方面、准确捕捉这种复杂性的挑战方面、他们自己参与概念建模的必要性方面、产生高质量的灵活实现所需工程的复杂性方面有一些了解，从而来面对这样的事实——系统不是魔术，而是富有逻辑的良好判断体系，它是工程，凝结着很多艰辛的劳动。

实际上，每一位数据建模者，不管他（或她）名义上的专业有何不同，不管学的是 Chen, Barker, Finkelstein, IDEF1X, IDEF1x（对象），UML, XML，还是 XYZ，都应该因为同样的理由来读读这本书。事实上，每个建模者、每个程序设计员、每个与信息或信息系统打交道的人都应该好好地读读这本书。

本书的另一个精髓在于，通过利用面向角色建模导出了语义表达的高标准。一旦论证出有可能严谨地以图解的方式来掌握精确的语义目标，我们就可以按照它们复制表达的能力来非常直接地评估所有其他的图解模型符号。Terry 比较了我所听到过的每一种建模符号（包括我刚刚提到的那些）。他对于自己就其他语言持批判态度感到有些抱歉，但我的看法是，他所做的比较是非常公

正、准确和客观的。Terry 甚至指出了其他那些符号的精髓所在，以及在整个过程中它们能在哪里及以何种方式得到有效的利用。

这些严谨、精确的语义模型还有着更为有趣的方面——它们可以转化成数据库。Terry 通过例子详细描述了从概念模型到逻辑模型、物理数据库设计、实现的整个转换。由此可以很容易地评价和比较包括关系数据库、面向对象数据库、对象 - 关系型数据库、描述型数据库等在内的不同数据库实现的可能性；同时，他还为星形模式和时空数据库提供了良好的衡量标准。由此，我们可以很直接地对在不久的将来数据库技术发展的趋势做出一个合理和现实的规划。

我想，这本书会永远珍藏在我的书架上。我将它视为在这个尚属年轻的、仅有 50 年历史的信息管理学科中为数不多的能称得上“经典”的作品之一。我相信，你也会像我一样发现这本书中蕴含着的实用性和启迪性。

序二

Robert Meersman 教授，博士
比利时布鲁塞尔 Vrije 大学计算机科学系
系统技术与应用研究实验室主任

我很高兴能再次向各位读者介绍 Halpin 博士继他的前一本书后新近推出的这本非常有意义的新作——《信息建模与关系数据库——从概念分析到逻辑设计》，相对于前一本书，它在许多地方都做了进一步的润色和更新。我将他的前一本书作为布鲁塞尔 Vrije 大学计算机课程的教材已有多年，相信这本书也同样会得到广大师生的认可。

与所有优秀的参考书籍一样，本书为 ORM 方法论的研究和应用提供了一个坚实、规范和权威的基础。这门学科所代表的科学技术，起源于 20 世纪 70 年代的“NIAM”，发展已有多年，却始终在文献方面未能有一本高质量的参考书籍。在某种意义上，这个缺陷已经制约了 ORM 及其工具在 DP 和 IT 领域以及数据库和信息系统研究领域的广泛推广。

我很荣幸，能够和其他同道一起投身到将这门学科应用到设在布鲁塞尔的 Nijssen 研究实验室中的工作上来。所有优秀的方法论都可以比做美酒，是长期酝酿的结果，需要经过许多的试验，需要许多实践者运用智慧对其进行多次的过滤提纯。可以说，只有一个真正有创造力的酒窖主人才能生产出真正意义上的好酒。Terry Halpin 将精湛的学术研究技巧与要为数据库从业者提供一本有用的工具书的愿望有机地结合起来，毫无疑问，他现在已经成为对象角色建模及相关领域里一流的专家。

读者可能已经发现了，诸如 EER 和 UML 等一些有竞争力而又不是很有活力的方法已经占据了主流数据库设计的“方法论市场”的一些份额。而近来的研究一直在不断证实着这样一种趋势：随着对 NIAM/ORM 语义的关注程度的逐渐提高，它很有可能成为计算机化语义仓库（也叫本体）新规格的更有资历的候选者，而其他方法也许注定会被代替。

最后，我衷心祝愿读者能够从 ORM 的学习中得到乐趣并将它很好地应用到实践中去。

序三

Gordon Everest, MIS 及 DBMS 教授
美国明尼苏达大学 Carlson 管理学院

能够为这本书作序,我感到很高兴,同时也非常荣幸。这又给了我一次机会来证明在数据建模领域中其实还存在着更好的方式,对象-角色建模(ORM)就是其中之一。我非常想激起大家认真学习 ORM 的兴趣,而本书正是你能够利用的最好资源。

数据建模是信息系统得以发展的基础,如果你不能“恰当地”设计数据库,那么所建立起来的系统就会像一间装满卡片的房子,在未来不可避免的修改、扩充、重新整合的过程中和提高质量的种种压力下最终崩溃。因此,需要一个模式来指导我们的数据建模,创造出清晰准确、能够代表用户论域、方便人与人相互交流理解的有效数据模型。

长期以来我一直都是“无记录型”数据建模的忠实信徒。1975 年从德国回来后,我开始在高级数据库课程中讲授“二元建模”,这门课被 MIS 专业的大多数 MBA 学生和博士生以及许多本科学生所选修。那些年,讲授 NIAM 非常困难,因为我们没有好的教材和可用的工具软件。直到 1990 年,才有了 Halpin 和 Nijssen 编写的这本书的早期版本,随后有了可用的、能在计算机上支持 ORM 的工具——InfoDesigner(即以后的 VisioModeler)。现在我们可以提供三个版本的高级数据库设计课程,每年可供 100 多名学生选修。学生们一致认为这本书是这一课程中最好的教材,它编排合理、简明易懂。

本书是每一个对数据建模感兴趣的人的必读之物,但是要注意:要理解 ORM,必须投入足够的时间和精力。这本书在上一版的基础上增加了更多的练习来帮助你提高数据建模的技巧,更好地理解 ORM。同传统的数据建模相比,ORM 的确与众不同。其他数据模型都建立在代表实体的记录之上——Chen 的实体-联系(ER)图以及它的许多变化形式、Teorey 的扩展型 ER、Finkelstein 的信息工程、Appleton 的 IDEF1X、Ullman 的语义数据模型(SDM)、Kroenke 的语义对象模型(SOM)、UML 的类图以及 Codd 在关系模型中对关系的论述。如今有许多 CASE 软件工具能够体现出这些数据建模模式。这种基于记录的模式包括三个有机组成部分:实体、属性和联系。在数据建模中,许多问题的根源在于把众多的属性集合成实体记录。规范化就是要通过测试来看看我们是否聚集了过多的属性,记录分解通常是修正违背范式情况的一种方法。如果能够充分地分解一个基于记录的设计,让每个记录都至多只有一个非键域,那么,就算是以 ORM 设计的方式解决问题了,可以使每一个记录都能和一个谓词相对应,这也就是联系。

规范化是数据建模的要害。怎样才能避免出现规范化呢?答案是利用 ORM 就可以了!问题的根源在于当我们把过多的属性塞进记录时导致了缺乏清晰的联系的界定,以致记录内部结构的定义或假设十分隐晦。ORM 可以促使你去思考并独立地定义你的研究领域中所涉及到的对象的相关联系和限制。

实际上,ORM 只建立在两个部分之上:对象和联系(这与一个句子中名词作主语或宾语,动词作谓语的概念相对应)。在 ORM 中实体和属性被视做对象(不要与面向对象技术的对象相混淆)。对象在与其他对象的联系中充当角色。实体或对象通过在与其它对象的联系中充当角色来拥有属性或描述符。当 ORM 模型能有效地代表需要建模的领域时,函数依赖就可以很清晰地定义出来,

“记录”(在关系表的图中)的产生可以自动与充分规范化的(达到 5NF)有保障的结果相匹配。这对数据建模者来讲称得上是个好消息。

ORM 并不能代替 ER 图或关系数据库设计。相反,它只能是它们的前一阶段。因此,我们把它叫做概念建模,它先于“逻辑”建模阶段,后者通常是在关系数据库结构中建立记录。实际上,记录格式更多地是考虑到系统的效率,而不是为了便于使用者理解。对记录(许多属性的集合及代表实体的标识符)不太成熟的认识事实上已经成为数据建模的一个障碍。

与其他基于记录的模式相比,无论是图还是语言描述,ORM 都是一个更加丰富、更加富于表现力的数据建模模式。它能够捕捉到并更为直接地代表更多的语义(例如多对多的联系、三元和更高次序的联系,以及混合的实体间约束),无需引进诸如交叉实体或外键等人为的或伪造的成分,从而避免了对使用数据模型的用户领域的专家们造成不必要的干扰。

不知以上我所讲的能否激发出你足够的兴趣来学习 ORM 数据建模的一些知识?如果你有志于从事 ORM 的研究或是对数据建模感兴趣,我建议你投入一些时间和精力来好好地读读这本书,你会感到物有所值的!你会欣赏 ORM 并通过应用 ORM 成为一个更出色的数据建模者。要开发有效的、可维护的信息系统需要好的数据模型,为此需要一个好的数据建模模式。ORM 可以让我们在最高的概念层次上进行数据库设计,并忽略那些不需要被用户领域的专家所关注的内容。我深切地希望看到越来越多的数据库设计者开始使用 ORM,我们所建立的系统和我们所生活的世界都会因此而受益的!和我一起开始探索之旅吧,相信你一定会喜欢的!

前 言

这是一本关于信息系统的书,重点讨论信息建模和关系数据库系统,主要适用于数据库专业人员、计算机科学和信息管理专业的学生。它也适用于那些想以人们容易理解且在计算机上容易实现的方式来阐述应用信息结构的人。此外,本书还提供了—个简单的概念框架来让读者理解数据库系统到底是什么,并对 SQL 进行了详尽的介绍。

该书主要研究对象—角色建模(ORM),这是一种依照对象和它们扮演的角色来观察世界的概念建模方法。ORM 起源于欧洲,在那里它以其他的名字,如 NIAM (natural-language information analysis method, 自然语言信息分析方法)或 FCO-IM (Fully Communication Oriented Information Modeling, 完全面向交流的信息建模)为人所知。本书所用的 ORM 版本是在 NIAM 基础上进行扩充的,并拥有微软公司的工具支持。

另外两个流行的信息建模符号表示法是实体—联系(ER)图和统一建模语言(UML)类图。对于概念信息分析,ORM 比 ER 图及 UML 方法有一些优势。例如,ORM 模型易于语言表示和成员化,以便于该领域专家的确认,应用域变化时,它们比较稳定,而且它们能够在图表中得到更多的业务规则。然而,ER 图和 UML 类图适于简明的总结。它们的结构更接近于最终的数据库实现,所以,也是很有价值的。因此,本书涵盖了用 ER 和 UML 进行数据建模的内容,并说明了如何从 ORM 模型中很容易地提取出 ER 和 UML 数据模型。

为了使内容更接近大多数对数据库感兴趣的读者,本书行文简洁易懂,并且尽可能地回避形式化的数学分析和计算。必要时,基本逻辑和集合理论中的相关概念将在使用它们之前进行讨论。书中大部分材料已在工业界和学术界的课程中经过了分类检验,并且基本的 ORM 方法已在高校成功地讲授多年。各部分内容已经模块化,因此,想要省略一些内容的教师可为他们的课程做出合适的选择。

第 1 章阐明了概念建模研究的动机,并简要地比较了 ORM, ER 和 UML 方法。对信息系统的历史和结构也进行了概述。

第 2 章讲解了结构背景知识,解释了信息系统的概念结构和开发框架,并且介绍了后面一些章节需要用到的关键概念,只有很少或没有数据库经验的读者应全部阅读。

第 3 章是基础。先对概念建模语言标准和 ORM 概念设计过程(CSDP)进行了概述,然后,讨论了 CSDP 的前三个步骤。第一步(依照基础事实用语言表示常见的例子)看起来可能很琐碎,但不应匆忙进行,因为它为模型提供了基础。本章的剩余部分涉及了事实类型的基本图形符号,然后讲解了如何将对象分类并识别能用算术方法推导出的信息。

第 4 章讲述在事实类型成员集上指定约束。最重要的一种约束(惟一约束)被详细说明。随后讨论了事实类型基本性质的检验。本章同时还介绍了概念层的连接和投影操作——这些操作的关系版本在后面的关系数据库中非常重要。

第 5 章涉及了强制的角色约束,包括对探测可以逻辑导出的信息的检验。并在某种程度上检验了引用模式。对于短期课程,一些较复杂的引用模式可以跳过。本章将在实例中复习已经讲过的 CSDP 步骤。

第 6 章涉及值、集合比较(子集、等价和互斥)及子类约束。6.6 节研究子类高级方面的内容。虽然这对于实践很重要,但是第一次阅读时可以跳过这部分内容。

第7章涉及概念模式设计过程的最后一步。考虑了较少的公共约束(如出现频率和环约束),并且对设计进行最终检验。7.3节至7.5节有点儿高深,短期课程可以跳过。

第8章讨论了实体-联系(ER)方法,以Chen最初的符号表示法开始,然后讨论了当前使用的三种最流行的符号表示法:Oracle公司支持的Barker ER符号表示法、信息工程符号表示法,以及实际上是ER和关系符号混合物的IDEF1X符号表示法,并且与ORM进行了比较。

第9章研究用于数据建模的UML类图的用法,包括与ORM详细的比较。ORM中的业务规则结构在UML中没有相应的图形表示,它们是通过使用用户定义的约束和注释来标识和获取的。

第10章描述如何在关系数据库中实现概念模型。前三节是理解概念模式如何映射到关系模式的基础。10.4节讨论高级映射方面的内容,在入门课程中可以省略。

第11章研究一些关系数据库的查询语言。11.1节涉及关系代数,虽然它不作为一种实际的查询语言使用,但关系代数对于理解基本的SQL支持的关系操作很重要。11.2节将数据的关系模型与一些关系数据库管理系统使用的数据模型进行了概括比较。11.3节至11.4节涉及SQL的主要特性,重点在SQL-89,SQL-92及SQL:1999标准,以及一些流行的变体。

第12章讨论在相同级别(概念上或逻辑上)如何及何时将一个模式转换到另一个模式。12.1节至12.4节研究了概念模式等价性和概念模式改造的方法。作为这个理论的一个应用,12.5节说明了通过在映射前进行概念转换以优化数据库设计的过程。12.6节完整地介绍了规范化理论。12.7节简要介绍了反规范化和低级优化。12.8节说明了概念优化在数据库再工程化中扮演的角色。12.4节、12.5节、12.7节和12.8节涉及高级特性,在短期课程中可以跳过。对于非常短的课程,整章都可以跳过。

第13章研究其他的建模问题、方法和趋势。涉及的主题包括数据仓库、概念查询语言、模式抽象机制、处理模型(如UML示例和活动图、数据流图)、后关系数据库(如对象数据库和对象-关系数据库)以及元建模。虽然这些主题重要且有趣,但它们在短期课程中可被省略。

本书采用了大量的图和示例。每一章以简短的概述开始,以本章涉及的要点小结结束,此外还有章节注释以补充介绍要点和进一步的参考资料。本书的主要特点之一是带有大量的练习题,这些练习题全都经过课堂检验并被仔细分级。全书最后列出了引用的全部参考书目,以及技术符号和ORM,ER及UML(只有类图)的术语表。

在线资源

本书大量的补充材料在出版商的网站(www.mkp.com/imrd/)上可以在线下载。其中有三个附录:附录A提供了计算机硬件和软件演变的概述;附录B讨论了两类用于从有意义的成员集中确定子类图的子类矩阵;附录C介绍了SQL的高级方面,集中于集合比较查询和组极值查询。

练习题的答案包含在两个文件中:一个是编号为奇数问题的答案,一个是编号为偶数问题的答案。编号为奇数问题的答案可以公开得到,编号为偶数问题的答案有密码保护,以便提供给授课者用来作为课堂讨论练习题。关于ORM的其他材料可以从作者的网站www.orm.net上获得。

图表的电子版,以及进一步的练习和相关教学材料包括在附送的教师指南中。授课者可以通过与www.mkp.com/academic/rep_locator.asp中列出的代表联系,来获得这本指南和练习答案的密码。

ORM 软件

ORM得到来自微软公司和其他公司的各种建模工具的支持。一个废弃的ORM建模工具VisioModeler可以从微软的MSDN网站免费获得。尽管该下载软件没有技术支持,并且该产品的数据库驱动支持有些过时,但它对于学习ORM仍很适合,并且它允许你创建ORM模型并映射到一

些数据库系统中。要获得这个下载软件（25 MB），请将你的浏览器指向 <http://download.microsoft.com/>，然后进行“VisioModeler”关键词搜索。

在写本书时，从微软的 Visio Enterprise 2000 产品中可以获得基本的 ORM 建模解决方案。这个软件比 VisioModeler 提供了更好的驱动支持，并且包括对视图和代码编辑的支持，但它显示 ORM 约束的能力非常有限。Visio 未来的版本将支持 ORM 图，但不支持映射。

一个基于 Visio 的 ORM 建模解决方案的完全改进版本出现在 Visual Studio.net 中。这个解决方案不仅能显示所有的 ORM 约束，还包括许多说明和映射逻辑数据库模型的改进方案。这个解决方案在 Visual Studio.net 的 beta 2 版本中首次出现。关于微软的 Visual Studio 的详细资料可以从 <http://msdn.microsoft.com/vstudio/> 上获得。

目 录

第1章 引言	1
1.1 信息建模	1
1.2 建模方法	4
1.3 一些背景知识	12
1.4 相关技术	15
1.5 小结	16
本章注释	16
第2章 信息层次及框架	18
2.1 四个信息层	18
2.2 概念层	21
2.3 从外部到概念再到关系	29
2.4 开发框架	34
2.5 小结	38
本章注释	38
第3章 概念建模：第1步	40
3.1 概念建模语言标准	40
3.2 ORM 概念模式设计过程	42
3.3 CSDP 第1步：从例子到基础事实	43
3.4 CSDP 第2步：画出事实类型，填充成员	56
3.5 CSDP 第3步：整理模式，注释基本推导	67
3.6 小结	74
本章注释	75
第4章 惟一性约束	78
4.1 CSDP 第4步：惟一性约束，元数检查	78
4.2 一元和二元的惟一性约束	79
4.3 较长事实类型上的惟一性约束	87
4.4 外部惟一性约束	92
4.5 键长度检查	99
4.6 投影连接检查	109
4.7 小结	113
本章注释	114

第5章 强制角色	115
5.1 CSDP第5步介绍	115
5.2 强制角色和可选角色	117
5.3 逻辑推导检验	126
5.4 引用模式	131
5.5 用例研究：光盘零售商	146
5.6 小结	149
本章注释	150
第6章 值、集合比较和子类约束	151
6.1 CSDP步骤6：值、集合和子类约束	151
6.2 基本的集合理论	151
6.3 值约束和独立对象	154
6.4 子集、相等和排斥约束	161
6.5 子类化	173
6.6 对象类型的一般化	189
6.7 小结	194
本章注释	196
第7章 其他约束和最后检验	197
7.1 CSDP第7步：其他约束和最后检验	197
7.2 出现频率	197
7.3 环约束	201
7.4 其他约束	210
7.5 最后检验	217
7.6 小结	221
本章注释	222
第8章 实体-联系建模	223
8.1 ER概述	223
8.2 Barker符号	224
8.3 信息工程符号	232
8.4 IDEFIX	236
8.5 从ORM映射到ER	244
8.6 小结	247
本章注释	247
第9章 用UML进行数据建模	249
9.1 引言	249
9.2 对象引用和单值属性	250
9.3 多值属性	254

9.4 关联	258
9.5 集合比较约束	263
9.6 子类化	269
9.7 其他约束和推导规则	272
9.8 从 ORM 到 UML 的映射	282
9.9 小结	284
本章注释	286
第 10 章 关系映射	287
10.1 实现概念模式	287
10.2 关系模式	288
10.3 关系映射过程	293
10.4 高级映射方面	314
10.5 小结	325
本章注释	326
第 11 章 关系语言	327
11.1 关系代数	327
11.2 关系数据库系统	346
11.3 SQL: 历史及结构概述	347
11.4 SQL: 标识符和数据类型	349
11.5 SQL: 选择列、行和排序	353
11.6 SQL: 连接	360
11.7 SQL: in, between, like 和 is null 操作符	368
11.8 SQL: 并和简单子查询	374
11.9 SQL: 数量操作符和包函数	382
11.10 SQL: 分组	388
11.11 SQL: 关联和存在子查询	394
11.12 SQL: 数据定义	399
11.13 SQL: 更新表数据	408
11.14 SQL: 安全和元数据	410
11.15 小结	410
本章注释	413
第 12 章 模式转换	414
12.1 模式等价和优化	414
12.2 谓词专门化和通用化	417
12.3 嵌套、互参考和平凡化	426
12.4 其他转换	440
12.5 概念模式优化	444
12.6 规范化	454