



注册电子商务工程师(CEBE)认证培训教材



电子商务 的数据管理技术

注册电子商务工程师认证培训教材编委会组织编写

徐保民 主 编

孙丽君 孟昭璟 副主编



中国水利水电出版社
www.waterpub.com.cn

注册电子商务工程师（CEBE）认证培训教材

电子商务的数据管理技术

徐保民 主编

孙丽君 孟昭璟 副主编

中国水利水电出版社

内 容 提 要

数据管理技术是实现电子商务信息处理的基础和条件,是电子商务的关键支撑技术之一。本书重点讲述数据管理技术的产生与发展、关系数据库的基本理论和设计理论、Web 数据库技术、电子商务中的数据库技术、数据库管理系统产品及电子商务系统的构造等内容。

本书编写本着从易到难,循序渐进,理论与实践并重的原则,力求通俗易懂,讲求实用性和先进性。

本书可作为高等院校电子商务、经济管理及其他相关专业的教材和对电子商务中数据管理技术感兴趣的读者参考、学习和使用。

本书所配电子教案可以从中国水利水电出版社网站上免费下载,网址为:
<http://www.waterpub.com.cn/softdown/>。

图书在版编目(CIP)数据

电子商务的数据管理技术 / 徐保民主编. —北京:中国水利水电出版社, 2005

(注册电子商务工程师(CEBE)认证培训教材)

ISBN 7-5084-3246-0

I. 电… II. 徐… III. 电子商务—数据管理—技术培训—教材
IV. ①F713.36 ②TP316

中国版本图书馆CIP数据核字(2005)第103510号

书 名	电子商务的数据管理技术
作 者	徐保民 主编 孙丽君 孟昭璟 副主编
出版 发行	中国水利水电出版社(北京市三里河路6号 100044) 网址: www.waterpub.com.cn E-mail: mchannel@263.net (万水) sales@waterpub.com.cn 电话: (010) 63202266 (总机)、68331835 (营销中心)、82562819 (万水)
经 售	全国各地新华书店和相关出版物销售网点
排 版	北京万水电子信息有限公司
印 刷	北京市天竺颖华印刷厂
规 格	787mm×1092mm 16开本 12.5印张 304千字
版 次	2005年9月第1版 2005年9月第1次印刷
印 数	0001—5000册
定 价	20.00元

凡购买我社图书,如有缺页、倒页、脱页的,本社营销中心负责调换

版权所有·侵权必究

编 委 会

主 任：郭增利

副主任：孙春亮 杨庆川 连卫民 司志刚 濮小金

委 员：（按姓氏笔画排序）

上官绪智 卫 琳 石 云 石 磊

吉永宏 劳帼龄 张景学 张景波

席宁华 席红霞 徐保民 常朝稳

鹤荣育

序

互联网，催生管理和技术革命的火种；它使管理进步、技术升级频频闪现。互联网，深入经济和社会发展的砥柱；它使经济增长、社会前进蒸蒸日上。互联网，使复杂变得简单，使枯燥变得生动，使遥远变得贴近；世界随之而变，生活随之而变。

互联网让高科技迅速转化为现实生产力，互联网让商务活动高效转化为电子化的执行程序。电子商务正在成为一切经济活动不可或缺的组成元素，实现着过去难以逾越的服务功能和服务手段，信息传递、资讯交流、市场开拓得以在刹那之间成功完成。

电子商务正在成为助推企业发展的核心力量，转变着以往粗放型经济高成本、低收益的产业格局。电子商务正在成为无可替代的时代风向标，超越国界，超越人的想象空间，改变着物理上的衡量尺度。

电子商务的地位和作用已经难以撼动，展望未来，其发展更会随着观念的进步、技术的成长，应用能力的提升而走向更高的顶点。

这是一项没有尽头的事业。

电子商务是一门高度复合的科学。电子商务是本土的，更是国际的；电子商务是技术的，更是管理的；电子商务是理论的，更是实践的。新的时代强烈呼唤着电子商务人才快速崛起，新的时代期待着电子商务人才队伍日益壮大。

2005年，中国的电子商务终于全面走向国际化和本土化整合的专业化发展道路。在中国商业联合会的支持下，中国将首次全面引进国际商务职业资格认证协会（ICPQA）与国际电子商务工程师协会（ICEBE）共同组织实施的专业技术和应用型培训考试项目——注册电子商务工程师考试（CEBE）。中国电子商务专门人才培养体系将由此而真正诞生。

CEBE 是国际电子商务人才的专业化认证体系，遵循国际化执行程序 and 标准，并采纳适用的本地化建议而设置的专业考试课程，它的引进将极大地提高我国电子商务人才的专业技能，进一步优化电子商务人才结构，加速中国电子商务人才与国际专业的接轨进程。

CEBE 教材和课程体系涵盖《注册电子商务工程师考核大纲》规定的全部内容。本套丛书包括《电子商务导论》、《电子商务的营销技术》、《电子商务的网络技术》、《电子商务的数据管理技术》、《电子商务的安全技术》、《电子商务的应用开发技术》等六大方面的专门技术，以及《电子商务案例分析》。

《电子商务导论》包括电子商务的基本概念和模型、电子商务技术基础、电子商务网站建设、电子商务环境、电子商务运营方式等内容。《电子商务的营销技术》包括电子商务营销的基本理论、网络市场调研、电子商务营销战略与计划、电子商务营销广告、网络销售、网络销售的评价要求等内容。《电子商务的网络技术》包含计算机网络基础、Internet 技术、数据通信基础、网络安全、电子商务金融网络接入方案和电子商务策划与实施等内容。《电子商务的数据管理技术》包括数据库的基本理论、面向对象的数据库、Web 数据

库基础、数据仓库、数据库技术与电子商务、数据库产品等内容。《电子商务的安全技术》包括电子商务安全的现状与趋势、信息加密技术与应用、数字签名技术与应用、TCP/IP 服务与 WWW 安全、防火墙的构造与选择、计算机病毒及其防治、系统评估准则与安全策略、计算机信息系统安全保护制度等内容。《电子商务的应用开发技术》包括电子商务工程及应用框架、HTTP 与超文本标记语言 HTML、客户端技术、服务器端开发技术、网络安全开发技术、XML 应用开发技术等内容。《电子商务案例分析》包括企业建网方案、数据库管理方案、与金融网络的接入方案、商品营销方案、信息安全方案、物流配送方案等。

CEBE 进入中国，受到了中国众多高校的支持，中国水利水电出版社联合解放军信息工程大学、北京理工大学、对外经济贸易大学、郑州大学、北京交通大学、山东大学、上海财经大学等高校电子商务和计算机网络专家对教材体系和认证体系进行了本地化的拓展和完善，以期为中国电子商务专门人才培养提供有效的本土化和国际化整合的认证通路，为 CEBE 培训和考试提供重要、充分的保障。

本系列教材体系有如下几方面的特点：

1. 针对性强。主要根据《注册电子商务工程师考核大纲》为注册电子商务工程师考试而编写。

2. 实用性强。丛书以技术为主线，突出实际应用，丛书的作者都是长期从事电子商务技术和计算机网络技术教学、研究和开发的专家，书中许多技术就是他们经验的总结，这对电子商务人才的培养具有重要意义。

3. 体系结构合理。针对人们认识问题的规律，强调面向应用，注重应用能力培养，层次清晰。

4. 适用广泛。由于是为电子商务技术人才培养而编写的丛书，所以这套丛书也适合各高等院校电子商务和计算机相关专业学生，以及社会在职人员学习和使用。

该教材体系和本套丛书将以统一规划、分批组织、陆续出版为原则。希望各位专家和同行及时对本套丛书给予指正，使其进一步完善，以形成适应中国电子商务发展需要的专业化目标。

中国国际商务职业资格认证管理办公室

2005 年 1 月

前 言

20 世纪 90 年代人类社会信息化进程的一个重大变化就是 Internet 的出现。现在 Internet 已经从单纯的学术科研网络向综合性商业网络发展,运用 Internet 进行电子商贸活动风靡全世界。我国从 1999 年开始,特别是 1999 年下半年,也掀起了电子商务的热潮。

电子商务是现实社会中商务活动的电子化实现方式,因此它的实现过程就是现实中商务活动的各个过程,包括交易前商品信息的浏览、定单的签署、支付货款、商品的配送和售后服务等过程。其中,在为用户提供商品信息时需要有商品信息库,而信息库的建立需要有数据库技术。同时,用户需要使用 Web 浏览技术浏览商品信息并完成商品订购。总之,电子商务的应用离不开网络技术和数据库技术的支持。而数据库技术在电子商务应用中起着数据管理的重要作用。

本书主要介绍电子商务中起着数据管理作用的数据库技术和 Web 数据库技术,目的是为读者全面了解数据库技术和 Web 数据库技术在电子商务中的应用打下坚实的基础。

本书共 7 章。第 1 章数据库技术概述,第 2 章关系数据库的基本理论,第 3 章关系数据库的设计理论,第 4 章电子商务与数据库技术,第 5 章 Web 数据库技术,第 6 章关系数据库管理系统简介。第 7 章构造电子商务系统。

参与本书编写的老师,多年来一直从事教学工作,积累了丰富的教学经验,书中的很多内容都是他们教学经验的总结。本书由徐保民主编,并负责统稿和定稿工作,孙丽君、孟昭璟担任副主编。各章编写分工如下:徐保民编写了第 1 章、第 2 章和第 3 章的内容,孙丽君编写了第 4 章和第 6 章的内容,孟昭璟编写了第 5 章和第 7 章的内容。本书在编写过程中,得到了连卫民教授、石磊教授、马志伟教授等的悉心指导,在此表示深深的感谢。

由于时间仓促及作者水平有限,书中难免有疏漏之处,欢迎广大读者批评指正。

作 者

2005 年 7 月

目 录

序

前言

第 1 章 数据库技术概述	1
1.1 信息与数据	1
1.1.1 信息	1
1.1.2 数据	2
1.1.3 信息与数据的关系	2
1.2 数据处理与数据管理	2
1.3 数据模型	3
1.3.1 数据描述	3
1.3.2 概念数据模型	4
1.3.3 逻辑数据模型	7
1.3.4 物理数据模型	12
1.4 数据库中的基本概念	12
1.4.1 数据库	12
1.4.2 数据库管理系统	12
1.4.3 数据库系统	13
1.4.4 数据库、数据库管理系统与数据库系统之间的相互关系	14
1.5 数据库系统的结构	14
1.5.1 数据库三级模式结构	14
1.5.2 数据库系统的体系结构	15
1.6 数据库技术的发展	16
1.6.1 面向对象数据库系统	17
1.6.2 分布式数据库系统	19
1.6.3 多媒体数据库系统	20
1.7 本章小结	21
习题	21
第 2 章 关系数据库的基本理论	27
2.1 关系数据库	27
2.1.1 关系模型的数据结构	27
2.1.2 关系模型的完整性	29
2.2 关系代数	31

2.2.1	传统的集合运算	32
2.2.2	专门的关系运算	34
2.3	结构化查询语言 SQL	41
2.3.1	SQL 语言概述	41
2.3.2	SQL 的数据定义	43
2.3.3	SQL 的数据操作	47
2.3.4	SQL 的数据控制	52
2.4	存储过程	53
2.5	本章小结	54
	习题	54
第 3 章	关系数据库的设计理论	64
3.1	关系模式规范化概述	64
3.2	函数依赖	65
3.2.1	函数依赖	65
3.2.2	多值依赖	66
3.2.3	键	67
3.3	规范化和范式	67
3.3.1	第一范式	67
3.3.2	第二范式	68
3.3.3	第三范式	70
3.3.4	Boyce Codd 范式	71
3.3.5	第四范式	71
3.3.6	关系规范化小结	73
3.4	数据库设计	73
3.4.1	数据库设计概念	73
3.4.2	数据库设计方法	73
3.4.3	数据库设计步骤	74
3.5	本章小结	78
	习题	79
第 4 章	电子商务与数据库技术	86
4.1	电子商务概述	86
4.1.1	电子商务的概念及特性	86
4.1.2	电子商务系统的组成	87
4.1.3	电子商务的产生与发展	88
4.2	电子商务中的数据库技术	91
4.2.1	数据资源管理技术	91
4.2.2	数据库技术	95

4.2.3 决策支持技术	98
4.3 本章小结	108
习题	109
第 5 章 Web 数据库技术	114
5.1 Web 概述	114
5.1.1 Web 的发展历史	114
5.1.2 Web 的特点	114
5.1.3 Web 基本元素	115
5.2 计算模型的演变	116
5.2.1 客户/服务器计算模型	116
5.2.2 面向 Web 的计算模型	117
5.3 Web 数据库	119
5.3.1 Web 数据库的内涵	119
5.3.2 Web 数据库的特点	119
5.3.3 Web 数据库应用系统	120
5.4 Web 数据访问技术	122
5.4.1 访问 Web 数据库的原理	122
5.4.2 访问 Web 数据库的方法	122
5.5 本章小结	126
习题	126
第 6 章 关系数据库管理系统简介	132
6.1 数据库管理系统概述	132
6.2 Oracle 数据库系统	132
6.2.1 Oracle 概述	132
6.2.2 Oracle 系统特点	133
6.2.3 Oracle 的数据仓库解决方案	134
6.3 Informix 数据库系统	134
6.3.1 Informix 概述	134
6.3.2 Informix 系统特点	135
6.3.3 Informix 的数据仓库解决方案	136
6.4 SQL Server 数据库系统	137
6.4.1 SQL Server 概述	137
6.4.2 SQL Server 系统特点	137
6.4.3 SQL Server 多版本支持	138
6.4.4 SQL Server 的数据仓库解决方案	139
6.5 本章小结	139
习题	139

第7章 构造电子商务系统	142
7.1 电子商务系统模式	142
7.2 电子商务系统设计	143
7.2.1 电子商务系统构造原则	143
7.2.2 电子商务系统设计步骤	144
7.3 Web 数据管理	146
7.3.1 数据库管理系统选择	146
7.3.2 核心数据处理	147
7.4 本章小结	147
习题	147
附录1 参考答案	149
附录2 国际注册电子商务工程师（CEBE）认证考核大纲	185
参考文献	189

第 1 章 数据库技术概述

随着科学技术的高速发展以及计算机在我国各行各业的普及, 计算机应用已从科学计算、过程控制逐步转移到非数值处理的各个领域。

作为计算机软件的一个重要分支, 数据库技术是进行非数值处理行之有效的方法。因此近年来被广泛地应用于国民经济、文化教育和计算机辅助设计等领域。

要想掌握好数据库技术, 必须弄清数据、信息、数据管理、数据库、数据库系统等专业术语的内涵及他们之间的关系, 了解数据模型和数据库系统的结构等内容。

本章就是要介绍一些数据库系统的基本概念和基础知识, 以便让读者对数据库有一个概括性的了解, 可以说本章是了解数据库技术的向导, 也是进一步学习数据库技术的基础。

1.1 信息与数据

在数据管理技术中, 最常用到的概念就是数据和信息。他们之间既存在着联系, 又有着不同的含义。

1.1.1 信息

对于每个人来说, 对“信息”一词并不陌生。但作为日常用语, “信息”常指音信、消息; 而作为科学技术用语, “信息”被理解为对预先不知道的事件或事物的报道或者指在观察中得到的数据、新闻和知识。那么究竟什么是信息呢?

信息就是新的、有用的事实和知识。具体地说, 信息是一种已经被加工为特定形式的数据, 这种数据形式对接收者来说是有意义的, 而且对当前和将来的决策具有明显的或实际的价值。

信息具有如下一些重要的特性:

(1) 普遍性。信息是事物运动的状态和状态变化的方式, 因此, 只要有事物的存在, 只要事物在不断地运动, 就会有他们运动的状态和状态变化的方式, 也就存在着信息, 所以信息是普遍存在的, 信息具有普遍性。

(2) 无限性。在整个宇宙时空中, 信息是无限的, 即使是在有限的空间中, 信息也是无限的。一切事物运动的状态和方式都是信息, 事物是无限多样的, 事物的发展变化更是无限的, 因而信息是无限的。

(3) 相对性。对于同一个事物, 不同的观察者所能获得的信息量可能不同。

(4) 传递性。信息可以在时间上或在空间中从一点传递到另一点。

(5) 变换性。信息是可变换的, 可以有不同载体用不同的方法来载荷。

(6) 有序性。信息可以用来消除系统的不定性, 增加系统的有序性。获得了信息, 就可以消除认识主体对于事物运动状态和状态变化方式的不定性。信息的这一性质是信息对人类具有特别重要的价值。

(7) 动态性。信息具有动态性质, 一切活的信息都随时间而变化, 因此, 信息也是有时效的。信息是事物运动的状态和状态变化的方式, 事物本身在不断发展变化, 因而信息也会随之变化。所以人们在获得信息之后, 并不能就此满足, 信息要及时发挥效用, 要不断补充和更新。

(8) 转化性。信息可以转化, 在一定的条件下, 信息可以转化为物质、能量。最主要的条件是信息必须被人们有效地利用。正确而有效地利用信息, 就可能在同样的条件下创造更多的物质财富和能量。

在信息社会, 信息可以划分为两大类。一类信息能够用数据或统一的结构加以表示, 称之为结构化数据, 如数字、符号; 而另一类信息无法用数字或统一的结构表示, 如文本、图像、声音、网页等, 称之为非结构化数据。结构化数据属于非结构化数据, 是非结构化数据的特例。特别是随着网络技术的飞快发展, 使得非结构化数据的数量日趋增大。

1.1.2 数据

数据是用来记录信息的、可以识别的符号, 是信息的载体、信息的具体表示形式。在日常生活中, 数据无所不在, 数字、文字、图表、图像、声音等都是数据。

确切地讲, 数据具有两方面的含义: 一方面数据内容是事物特性的反映或描述; 另一方面数据是存储在某一媒体上符号的集合。由于事物特性的描述必须借助一定的符号即数据形式。而采用什么符号完全是一种人为约定, 因此数据可以有很多表现形式, 如文字、图形、图像及声音等。显然, 数据是需要解释的, 也就是要对数据的含义或语义进行解释。因为, 数据的形式即符号本身并不能完全表达其内涵。

1.1.3 信息与数据的关系

在现实生活中, 数据与信息两个术语常常被混淆, 但是它们之间还是有差别的。数据是信息的符号表示或称为载体; 信息则是数据的内涵, 是对数据语义的解释。即信息必然是数据, 但数据未必是信息, 信息仅是数据的一个子集, 有用的数据才成为信息。尽管数据和信息存在差别, 但在实际工作中, 二者常被不加区别地使用。

1.2 数据处理与数据管理

数据处理是指将数据进行加工转换成信息的过程。其目的是从某些已知的数据出发, 根据事物之间的固有联系和运动规律, 通过分析、归纳、演绎推导等手段加工出一些新的数据, 这些新的数据又表示了对人们有价值和有意义的新的信息。

数据管理是指数据的收集、整理、组织、存储、维护、检索、传送等操作, 这些操作是数据处理业务的基本环节, 而且是任何数据处理业务中必不可少的共有部分。数据管理的目的是由系统对数据进行高效、方便、有效地管理, 尽量减少数据管理人员的负担。

所以, 数据处理与数据管理是相互联系的, 数据管理技术的好坏, 将直接影响到数据处理的效率。

1.3 数据模型

数据模型是描述要管理的数据的逻辑结构、使用操作要求,以及完整性的形式方法。即数据模型是用来描述数据的一组概念和定义。一般来说,数据的描述包括两个方面:

- (1) 数据的静态性:包括数据的基本结构、数据间的联系和数据中的约束。
- (2) 数据的动态性:指定义在数据上的操作。

在数据库中,针对不同的使用对象和应用目的,采用多级数据模型,一般分为三级。一是独立与计算机实现的概念数据模型;二是直接面向计算机系统实现的、描述数据库中数据逻辑结构的逻辑数据模型;三是物理数据模型。

1.3.1 数据描述

在数据处理中对数据的描述将涉及到不同的范畴。具体讲,从客观事物到存放于计算机中的数据,这样一个加工过程可划分为三个世界:现实世界、信息世界和机器世界,即数据的三个范畴。

1. 现实世界

现实世界就是存在于人脑之外的客观存在的事物及其相互联系。事物之间的联系是由事物本身的特性决定的。

例如学校的教学系统中存在如下客观事物:教师、学生、课程;他们之间的联系为:教师为学生授课,学生选修课程并取得成绩。这就是现实世界,是数据库设计者接触到的最原始的数据。

2. 信息世界

信息世界,也称为概念世界,是现实世界在人们头脑中的反映,是对客观事物及其联系的一种抽象描述。信息世界是现实世界到机器世界必然经过的中间层次。在信息世界中,常涉及到的概念和术语包括:

(1) 实体。实体就是指现实世界中客观存在并且可以相互区别的对象。实体可以是具体的事物,如学生王五、数学课;也可以是抽象的概念或联系,如一门课程、学生王五选修了哪些课程等。

(2) 实体集。同型实体的集合称为实体集,如全体学生。注意区分实体集的“型”与“值”的概念,如每个学生是全体学生实体“型”的一个具体“值”。

(3) 属性。属性就是对实体所具有的某一特性的描述。通常,一个实体可以由若干个属性来刻画。如学生实体具有学号、姓名、性别、出生年月、入学时间等属性。

(4) 码。能够惟一标识一个实体的属性称为码。如学生的属性学号能够惟一确定一个学生,因此学号可以作为学生实体的码。

(5) 域。域就是属性的取值范围,如性别的域为(男,女)。

(6) 实体类型。具有相同属性的实体必然具有共同的特征和性质。所谓实体类型就是对具有相同属性的同类实体用实体名和属性名集合的抽象,即将具有共性的一类实体抽象为实体类型。如所有的男生具有共同的特征和性质,就是一个实体类型。

(7) 联系。联系是指实体与实体之间以及实体与组成它的各属性间的关系。通过联系就

可以用一个实体的信息来查找另一个实体的信息。

3. 机器世界

机器世界,也称为数据世界,是对信息世界中的有关信息进行加工、编码及格式化等具体处理,然后以数据形式存储在计算机内。可以说,机器世界中的数据既能展示信息世界,同时又便于用计算机进行处理。在机器世界中,常涉及到的概念和术语包括:

(1) 字段。字段就是标记实体属性的可以命名的最小单位,也称为数据项。字段命名通常与信息世界中的实体属性名相同。

几个字段可以组合在一起构成组合字段。如“日期”可以由年、月、日三个字段组合而成。组合字段也有自己的名字,可以作为一个整体看待。

(2) 记录。若干相关联的字段的集合构成记录。记录是应用程序输入、输出的逻辑单位。对于大多数据库系统来讲,记录是处理和存储信息的基本单位。通常用一条记录描述一个实体。构成该记录的字段表示实体的若干属性。如实体“学生”的一组数据(9103, 王五, 男, 计算机)就是一条记录。其中学号、姓名、性别及专业均为字段。

记录有“型”和“值”的区别。“型”是同类记录的框架;“值”是记录反映实体的内容。

(3) 文件。同一给定类型记录的集合构成一个文件。文件对应于信息世界的实体集,所以它又可定义为描述一个实体集的所有记录集。如所有的学生记录组成一个学生文件。

(4) 数据库。数据库是比文件更大的数据组织形式。数据库是指具有特定联系的多种类型的记录的集合。数据库由不同文件构成,这些文件之间存在某种联系。

(5) 关键字。关键字是能够惟一标识文件中的每一条记录的字段或字段集,它与信息世界中的码相对应,如学生的学号可以作为学生记录的关键字。

现实世界、概念世界和机器世界三者之间的关系如图 1-1 所示。需要说明的是,从现实世界到信息世界不是简单的数据描述,而是从现实世界中抽象出适合数据库技术处理的数据即实体,同时要求这些实体能够很好地反映客观世界中的事物;从信息世界到机器世界也不是简单的数据对应存储,而是要设计数据的逻辑结构和物理存储结构。所谓数据的逻辑结构是指程序员或用户用以操作的数据形式,是抽象的概念化数据;所谓数据的物理结构是实际存储在存储设备上的数据。

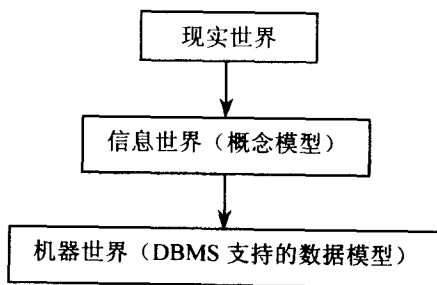


图 1-1 三个世界的关系

1.3.2 概念数据模型

概念数据模型,也称为概念模型,是指为正确、直观地反映客观事物及其联系,按用户观点对信息世界所建立的一个抽象模型。该模型独立于计算机系统的模型,完全不涉及信息在

系统中的表示,只是用来描述某个特定组织所关心的信息结构。实际上,概念模型是现实世界到机器世界的一个中间层次,是用户与数据库设计人员之间进行交流的语言,因此概念数据模型应该:

- (1) 具有较强的语义表达能力,能够方便、直接地表达应用中的各种语义知识。
- (2) 由于用户一般缺乏计算机知识,因此概念模型应当简单、清晰、易于理解。
- (3) 独立于计算机系统的模型,完全不涉及信息在计算机中的表示。
- (4) 独立于任何 DBMS,但可以转换为计算机上任一 DBMS 所支持的特定数据模型。

1. 实体与属性

我们已经知道,现实世界中任何客观存在的事物均可以是信息世界中的实体,这个事物可以是人,也可以是物;可以指实际的东西,也可以指概念性的东西。另一方面,信息世界中的任何一个实体都具有若干用于描述其特性的属性。

实体与属性的区分是相对的。在某个问题中可以看作实体的事物,在另一问题中可能作为属性存在。如考察交通工具的性能时,汽车、火车、马车等都是属性的身份出现的,而汽车本身又可以作为一个实体来考察它的各种属性。

现实世界中的事物之间或事物内部存在的联系,在信息世界中表现为实体集内部和实体集之间的联系。实体集之间的联系体现的是事物之间的语义关系。如“学生”实体和“课程”实体之间的联系体现的是一种“选修”联系。

必须强调的是,有时联系本身也有属性,这类属性不属于任一实体,只能属于联系。如联系“选修”还可以有成绩这一属性。

所以,作为对现实世界进行抽象描述的概念数据世界,除需对信息世界中的实体进行描述外,还应对其属性进行描述。

2. 实体间的联系

依据现实世界中事物之间存在的各种可能联系,可以将两个实体集之间的联系分为如下三类:

(1) 一对一联系。实体集 A 中任意一个实体至多对应实体集 B 中的一个实体,反之实体集 B 中的任意一个实体至多对应实体集 A 中的一个实体,则称实体集 A 与实体集 B 之间具有一对一联系。一对一联系记为 1:1。如一个人只有一种性别,一个人与性别之间为一对一的联系。

注意: 1:1 联系不一定都是一一对应的关系,可能存在着无对应。

(2) 一对多联系。实体集 A 中至少有一个实体对应实体集 B 中的 n 个实体 ($n \geq 1$),反之实体集 B 中的任意多个实体至多对应实体集 A 中的一个实体,则称实体集 A 与实体集 B 之间存在一对多联系。一对多联系记为 1:n。如一个父亲有多名子女,而一个子女只能有一个父亲,父亲与子女属于一对多的联系。

(3) 多对多联系。实体集 A 中的任意实体至少有一个实体对应实体集 B 中的 n 个实体 ($n \geq 1$),反之实体集 B 中的任意实体至少对应实体集 A 中的 m 个实体 ($m \geq 1$),则称实体集 A 与实体集 B 之间具有多对多联系。多对多联系记为 m:n。如一个学生可以选修多门课程,一门课程可由多个学生选修,学生和课程间存在多对多的联系。

实际上,一对一联系是一对多联系的特例,而一对多联系又是多对多联系的特例。实体集之间的这种一对一、一对多、多对多联系不仅存在于两个实体型之间,也存在于两个以上的实体集之间。同一个实体集内的各实体之间也可以存在一对一、一对多、多对多的联系,其含义

与实体集之间的联系类似。

3. 实体联系方法

由于直接对现实世界中的事物按具体的数据模型进行组织,设计工作不仅复杂,而且设计效果很难达到理想。因此需要一种方法能够对现实世界的信息结构进行描述。目前,最常用的方法是由美籍华人陈平山于1976年提出的实体—联系方法(Entity-Relationship Approach),简称E-R方法。应用E-R方法建立的概念模型称之为E-R模型,它是按用户的观点来描述现实世界的信息结构。

从本质上来讲,E-R方法是一种信息结构的图示法,E-R模型就是一种图示模型,所以E-R模型也称为E-R图。E-R模型通常由如下三个要素构成:

- (1) 实体:表示客观事物。
- (2) 属性:表示客观事物的特征(属性)。
- (3) 联系:客观事物之间的联系。

在绘制E-R模型时,通常使用如图1-2所示的图形符号来表示E-R模型中的三个要素。



图 1-2 E-R 图的基本图素

其中:

- (1) 矩形:表示实体,矩形框内写明实体名。
- (2) 椭圆:表示实体的属性,并用无向边将其与相应的实体连接起来。
- (3) 菱形:表示实体间的联系,菱形框内写明联系名,并用无向边分别与有关实体连接起来,同时在无向线段上标注联系的类型,即1:1、1:n或m:n。

注意:联系本身也可以是实体集,也可以有属性。如果一个联系具有属性,则这些属性也要用无向边与该联系连接起来。

由于E-R模型是在客观事物或系统的基础上形成的,在某种程度上反映了客观现实,反映了用户的需求,因此E-R模型具有客观性。但E-R模型又不等同于客观事物的本身,它往往反映事物的某一方面。至于选取哪个方面或哪些属性,如何表达则决定于观察者本身的目的和状态,从这个意义上说,E-R模型又具有主观性。所以,E-R模型具有客观性和主观性两重含义。但无论如何,E-R模型应该能够充分反映用户需求,并要得到用户的认可才能确定下来。

设计E-R模型的过程基本上是两大步:

- (1) 确定实体集及属性。针对特定用户的应用,首先确定哪些是实体集,有多少个实体集。其次确定各个实体集的属性。
- (2) 确定实体集间的联系。确定实体集之间存在什么联系及联系的属性。

数据库设计者的任务就是要把现实世界中的数据以及数据间的联系抽象出来,用“实体”与“联系”来表示。

另外,数据库设计者应注意,E-R模型应该充分反映用户需求,E-R模型要得到用户的认可才能确定下来。

例如,在产品供应数据库中,厂商数据以及产品数据是可被区分的对象,所以它们在数据