

中国计算机学会学术著作丛书

关系数据库理论

马 垣 著

清华大学出版社
广西科学技术出版社

(京)新登字 158 号

内 容 简 介

数据库技术是现代计算机技术中发展最快、应用最广泛的核心技术。关系数据库理论的发展,为数据库技术奠定了理论基础。在本书中,作者完整系统地介绍了关系数据库的概念和理论,并对近年来该领域的最新研究成果进行了较全面的总结。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

书 名: 关系数据库理论

作 者: 马垣

出版者: 清华大学出版社(北京清华大学校内,邮编 100084)

<http://www.tup.tsinghua.edu.cn>

印刷者: 昌平环球印刷厂

发行者: 新华书店总店北京发行所

开 本: 787×1092 1/16 **印张:** 25 **字数:** 600 千字

版 次: 1999 年 4 月第 1 版 1999 年 4 月第 1 次印刷

书 号: ISBN 7-302-03382-X/TP·1831

印 数: 0001~3000

定 价: 33.00 元

清华大学出版社 广西科学技术出版社
计算机学术著作出版基金

评 审 委 员 会

主 任 委 员 张效祥

副主任委员 汪成为 唐泽圣

委 员 王鼎兴 杨芙清 李三立 施伯乐 徐家福

夏培肃 董韫美 黄 健 焦金生

出 版 说 明

近年来,随着微电子和计算机技术渗透到各个技术领域,人类正在步入一个技术迅猛发展的新时期。这个新时期的主要标志是计算机和信息处理的广泛应用。计算机在改造传统产业,实现管理自动化,促进新兴产业的发展等方面都起着重要作用,它在现代化建设中的战略地位愈来愈明显。计算机科学与其它学科的交叉又产生了许多新学科,推动着科学技术向更广阔的领域发展,正在对人类社会产生深远的影响。

科学技术是第一生产力。计算机科学技术是我国高科技领域的一个重要方面。为了推动我国计算机科学及产业的发展,促进学术交流,使科研成果尽快转化为生产力,清华大学出版社与广西科学技术出版社联合设立了“计算机学术著作基金”,旨在支持和鼓励科技人员,撰写高水平的学术著作,以反映和推广我国在这一领域的最新成果。

计算机学术著作出版基金资助出版的著作范围包括:有重要理论价值或重要应用价值的学术专著;计算机学科前沿探索的论著;推动计算机技术及产业发展的专著;与计算机有关的交叉学科的论著;有较大应用价值的工具书;世界名著的优秀翻译作品。凡经作者本人申请,计算机学术著作出版基金评审委员会评审通过的著作,将由该基金资助出版,出版社将努力做好出版工作。

基金还支持两社列选的国家高科技重点图书和国家教委重点图书规划中计算机学科领域的学术著作的出版。为了做好选题工作,出版社特邀请“中国计算机学会”、“中国中文信息学会”帮助做好组织有关学术著作丛书的列选工作。

热诚希望得到广大计算机界同仁的支持和帮助。

清华大学出版社
广西科学技术出版社

计算机学术著作出版基金办公室

1992年4月

序 言

计算机是当代发展最为迅猛的科学技术,其应用几乎已深入到人类社会活动和生活的一切领域,大大提高了社会生产力,引起了经济结构、社会结构和生活方式的深刻变化和变革,是最为活跃的生产力之一。计算机本身在国际范围内已成为年产值达 2500 亿美元的巨大产业,国际竞争异常剧烈,预计到本世纪末将发展为世界第一大产业。计算机科技具有极大的综合性质与众多科学技术相交叉而反过来又渗入更多的科学技术,促进它们的发展。计算机科技内容十分丰富,学科分支生长尤为迅速,日新月异,层出不穷。因此在我国计算机科技尚比较落后的情况下,加强计算机科技的传播实为当务之急。

中国计算机学会一直把出版图书刊物作为学术活动的重要内容之一。我国计算机专家学者通过科学实践,做出了大量成果,积累了丰富经验与学识。他们有撰写著作的很大积极性,但相当时期以来计算机学术著作由于印数不多,出版往往遇到不少困难,专业性越强越有深度的著作,出版难度越大。最近清华大学出版社与广西科学技术出版社为促进我国计算机科学技术及产业的发展,推动计算机科技著作的出版工作,特设立“计算机学术著作出版基金”,以支持我国计算机科技工作者撰写高水平的学术著作,并将资助出版的著作列为中国计算机学会的学术著作丛书。我们十分重视这件事,并已把它列为学会本届理事会的工作要点之一。我们希望这一系列丛书能对传播学术成果、交流学术思想、促进科技转化为生产力起到良好作用,能对我国计算机科技发展具有有益的导向意义,也希望我国广大学会会员和计算机科技工作者,包括海外工作和学习的神州学人们能积极投稿,出好这一系列丛书。

中国计算机学会

1992 年 4 月 20 日

序

计算机技术在现代技术领域发展中最为迅速,应用最为广泛。数据库技术又是计算机技术中发展最快、应用最广泛的技术之一。当前,数据库技术已成为现代计算机信息系统和应用系统开发的核心技术,数据库已成为计算机信息系统和应用系统的组成核心。数据库技术更是未来“信息高速公路”的支撑技术之一。

数据库技术及其应用最早出现于 20 世纪 60 年代,我国对数据库的研究始于 70 年代中期。由于当时技术条件的决定,早期应用大多使用基于层次模型和网状模型的数据库管理系统。70 年代,E. F. Codd 提出数据库的关系模型,并对关系数据理论做出了创造性的工作和贡献,从而奠定了数据库技术的理论基础,使数据库从技术上升到理论。80 年代以来,关系数据库技术和理论得到长足的发展和广泛而有效的应用,出现了一批基于关系数据模型的商品化的数据库管理软件,它们从实验性系统发展成实用化、商品化的系统,在功能和性能上得到了极大的发展和提高,并形成了 SQL 语言的国际标准。实践证明,关系数据理论的研究成果不仅具有理论意义,更对数据库技术的应用具有指导意义。因此得到了数据库理论和应用人员的普遍重视,并引起不少从事数据库研究的人们对关系数据理论的兴趣。多年来所得到的研究成果从多方面丰富了关系数据理论的内容,获得了丰硕成果。本书正是对这些研究成果的较集中和较系统的反映和总结,也是作者长期从事关系数据理论学习和研究的结晶。

伴随着关系数据理论的发展,计算机技术、数据库技术以及与之相关的技术都同样得到发展。自 80 年代以来,数据库技术在技术发展和应用需求的双重影响下,发生了深刻且广泛的变化,这种变化主要表现在:一方面是多种技术与数据库技术的结合出现了数据库的多种新技术;另一方面是多种应用与数据库技术的融合出现了数据库的多种新应用。多种技术与多种应用的出现极大地丰富了数据库的新形式、新内容,从而在传统数据库的基础上形成并发展成数据库的大家族。

多种技术与数据库技术结合的典型是分布式数据库、并行数据库、多媒体数据库和面向对象数据库等技术的出现和形成。分布式数据库在数据库研究领域中有较长的历史,并出现过一批支持分布数据管理的系统,在国内也有一些实验性系统,但总的来说都还未能达到预期和理想的效果。而当今流行的 Client/Server 结构的数据库技术广义地理解也是一类分布式数据库技术。Client/Server 结构的数据库已成为当前的主流,而流行的商用系统仍大多是基于关系模型。

面向对象数据库是面向对象技术在数据库领域中的应用。面向对象概念起源于程序设计语言,更源于人们认识客观世界的一种方法。面向对象方法以客观世界中稳定存在的元素即对象为基础,采用软件技术来定义和刻画对象以及对象间的关系和联系。面向对象数据库的重要特点是它具有很强的描述客观世界中的复杂对象的能力,以及高效率开发应用和实现程序复用的能力。它是一类能有效地描述以多种媒体为表示形式并具有复杂结构应用的较理想的数据管理技术。面向对象数据库技术和系统的出现,曾经引起热烈的讨论,讨论的

主要内容之一是面向对象技术与关系技术间的关系,这涉及到二者的特点和表现能力以及关系数据库技术的局限和关系数据库系统是否会被面向对象系统所取代等问题,当然这种讨论也会影响关系数据理论的发展和前景。而从大家族的角度来看,它们将会并存和并行发展,一方面,这是由它们各自的特点和用途所决定,客观需求的差别会决定人们对系统的选择;另一方面,如果从数据库系统的实现技术和理论支持来分析,那就是关系数据库发展中所积累的知识和经验将可能在数据库技术的未来发展中得到永恒的借鉴。从当前数据库系统的技术实现途径中已经可以看出这种趋势。当前实现数据库管理的方式和技术大致可分为三类:RDB,即基于关系模型或称纯关系;OODB,即基于面向对象模型或称纯对象;ORDB,即基于对象-关系模型,它是二者的混合。后一种方式可视为对象技术与关系技术的结合或融合。人们预测 ORDB 将会是未来数据库技术发展的主流方向,因为它在一定程度上弥补了关系系统和对象系统的局限,并填补了二者之间的鸿沟。从历史的经验和哲学的观点来看,这种趋势可能是必然和自然的,人们总是在已经取得的成果上去获得新的发展。

多种应用与数据库技术的结合的内容十分丰富,产生了数据库大家族中的赫赫成员:工程数据库、演绎数据库、时态数据库、统计数据库、空间数据库、科学数据库、文献数据库等。它们大多继承了传统数据库,特别是关系数据库的理论和技術,但其功能和应用又有了新的发展。面向对象技术的引入也必然会给某些新的应用以新的技术支持,诸如工程数据库、空间数据库、科学数据库等具有复杂管理对象的应用,它们需要支持多种媒体形式的数据,需要定义复杂对象和对象间的复杂关系,这些都可能是传统的关系模型所难以表示的。数据库技术支持特殊应用的早期方式大多是将应用要求实现在数据库系统的应用层,并由用户来完成,而当前的应用需求大多却可以由数据库的内部功能来支持,实现了相关技术与数据库技术的融合,但就总体来看,其基础大多仍然是关系数据库技术与系统。

综上所述,对于璀璨的数据库大家族,基于关系数据理论的关系数据库技术现在仍然是它们的主要技术基础,关系数据理论和技术仍将直接或间接地支持着数据库技术和应用的发展,因此对其研究和发展在新形势下仍然是十分有意义的,本书的出版将为这种需要提供最好的支持。

在本书中,作者介绍了关系数据库的概念、理论和研究新成果,是对关系数据理论完整、系统的总结。其特点是作者收集了所能收集到的著作和文献,且加以系统化和条理化,并融入了作者的心得和成果。准确地说,作者应该是一位关系数据理论的业余爱好者和业余作者,他本人是冶金部鞍山黑色冶金研究设计院的高级工程师,只有在繁忙的工作之余才能从事对关系数据理论的研究。本书的完成实为难能可贵。

在即将完成本序的时候,我对作者的工作表示钦佩并对作者的精神表示敬意。感谢作者对我国数据库理论工作的建设所做出的贡献,并十分高兴地将本书介绍给读者。

中国科学院研究生院 罗晓沛

1998年3月于北京

目 录

第一章 关系模型	1
1.1 基本定义	1
1.2 关系运算	2
1.2.1 对元组的运算	2
1.2.2 关系代数	3
1.2.3 关系演算	7
第二章 函数依赖	10
2.1 问题的提起	10
2.2 函数依赖	11
2.2.1 函数依赖的定义	11
2.2.2 函数依赖模式	11
2.2.3 函数依赖的公理系统	12
2.3 第三范式及 BC 范式	15
2.4 不好的关系模式弊病产生的原因	17
第三章 函数依赖模式分解	19
3.1 模式分解的定义	19
3.2 无损连接的分解	21
3.2.1 定义	21
3.2.2 判定算法	22
3.2.3 算法的证明	24
3.3 无损依赖的分解	25
3.3.1 定义	25
3.3.2 判定方法	25
3.3.3 最小函数依赖集	25
3.4 转化为 3NF 的分解	27
3.5 转化为 BCNF 的分解	28
第四章 多值依赖	31
4.1 多值依赖的定义	31
4.1.1 问题的提起	31
4.1.2 多值依赖的定义	32
4.2 多值依赖的公理系统	33
4.2.1 公理系统	33

4.2.2	有效性的证明	33
4.2.3	完备性的证明	34
4.3	多值依赖的一些特性	38
4.3.1	多值依赖的无损连接特性	38
4.3.2	多值依赖的其他特性	38
4.4	多值依赖的依赖基	39
4.4.1	基本定理	39
4.4.2	求多值依赖依赖基的算法	40
4.4.3	算法正确性的证明	44
4.5	第四范式	46
4.6	嵌入的多值依赖与子集的依赖	49
4.6.1	嵌入的多值依赖的定义	49
4.6.2	嵌入的多值依赖的公理系统	50
4.6.3	子集依赖	51
4.6.4	Z-子集依赖的完备公理系统	53
4.6.5	从Z-嵌入的多值依赖集合推导一般的嵌入的多值依赖	56
4.6.6	嵌入的多值依赖不存在完备公理系统的证明	58
第五章	连接依赖与广义依赖	61
5.1	问题的提起	61
5.2	连接依赖的定义	62
5.3	完全连接依赖的有效公理系统	63
5.3.1	公理系统	63
5.3.2	有效性的证明	63
5.3.3	有向无回路图	64
5.4	广义依赖	69
5.4.1	等值产生依赖	
5.4.2	无组产生依赖	70
5.4.3	广义依赖模式	73
5.5	追赶算法	74
5.6	全连接依赖的完备公理系统	79
5.6.1	成功追赶的 DAG	79
5.6.2	增广连接依赖	82
5.6.3	完全连接依赖的完备公理系统	
5.7	第五范式	91
第六章	泛关系	94
6.1	泛关系的基本概念	94
6.1.1	物理导航与逻辑导航	94
6.6.2	泛关系的谓词定义	95

6.2	泛关系的连接依赖表征	97
6.3	泛关系中的空值	99
6.4	全投影与泛例及效模式	100
6.5	代表泛例	103
6.5.1	元组的淹没与关系的淹没	103
6.5.2	代表泛例的定义	103
6.5.3	求代表泛例的算法	104
6.6	泛关系的查询解释	112
6.6.1	窗口函数	109
6.6.2	唯一性模式与扩展连接	109
6.6.3	语义结构的窗口函数	112
第七章	无回路数据库	116
7.1	泛关系上的查询表达式	116
7.2	泛关系上查询可能的二义性	120
7.2.1	一个实例	120
7.2.2	克服二义性的一些方法	122
7.3	数据库模式的超图表示	123
7.3.1	超图	123
7.3.2	二义性与回路	125
7.4	α 回路	126
7.4.1	部分边集、关节、块与 α 回路的定义	126
7.4.2	α 无回路的等价特性	127
7.4.3	α 无回路十二种特性等价的证明	133
7.4.4	α 无回路数据库的实例	151
7.5	β 回路	153
7.5.1	β 回路的定义	153
7.5.2	β 回路的等价特性	154
7.5.3	β 回路五种特性等价的证明	156
7.5.4	β 无回路数据库的实例	159
7.6	Υ 回路	160
7.6.1	Υ 回路的定义	160
7.6.2	Υ 回路的等价定义	161
7.6.3	Υ 回路的四个定义等价的证明	161
7.6.4	Υ 无回路数据库的实例	163
7.7	Υ 无回路数据库泛关系查询无二义性	165
7.7.1	泛关系查询无二义性的形式化定义	165
7.7.2	几个定义	165
7.7.3	Υ 无回路泛关系查询无二义性的证明	166
7.8	各种无回路的识别与设计	169

第八章 数据库超图的闭包	178
8.1 数据库超图	178
8.1.1 有向一无向超图	178
8.1.2 超图的等价	179
8.2 超图的闭包	185
8.2.1 普通图闭包的推广	185
8.2.2 L-闭包	186
8.2.3 U-闭包	195
8.3 e-独立超图的闭包	202
8.3.1 独立超图	202
8.3.2 e-独立超图	204
8.3.3 e-独立超图的识别算法	205
8.4 e-无回路超图的闭包	207
8.4.1 无回路超图	207
8.4.2 e-无回路超图	212
8.4.3 e-无回路超图的识别算法	213
第九章 元组序列与字典序索引	214
9.1 元组序列	214
9.1.1 定义与基本运算	214
9.1.2 元组序列的右商	216
9.1.3 元组序列适合连接依赖的条件	218
9.2 索引	219
9.2.1 字典序索引的定义	219
9.2.2 索引的蕴含	220
9.2.3 索引蕴含推导的公理系统	220
9.2.4 公理有效性的证明	221
9.2.5 公理完备性的证明	222
9.2.6 适合给定索引集合的元组序列	223
9.3 索引与函数依赖	224
9.3.1 索引与函数依赖联合推导公理系统	224
9.3.2 联合公理系统有效性完备性证明	226
9.4 索引依赖	229
9.4.1 问题的提起	229
9.4.2 索引依赖的普遍性	231
9.4.3 索引依赖的形式化定义	232
9.4.4 索引依赖的背景异常	236
9.4.5 正则背景	238
9.4.6 索引依赖一般背景正则化	241

第十章 模糊关系	243
10.1 模糊关系模型.....	243
10.1.1 模糊集合.....	243
10.1.2 模糊关系定义.....	244
10.1.3 1型模糊关系实例.....	245
10.1.4 2型模糊关系实例.....	246
10.2 模糊关系运算.....	249
10.2.1 投影运算.....	249
10.2.2 延伸运算.....	249
10.2.3 自然连接运算.....	250
10.3 模糊整体约束.....	251
10.3.1 域依赖与数据依赖.....	251
10.3.2 模糊运算的传递原理.....	252
第十一章 模糊函数依赖	254
11.1 模糊域中的域值相等.....	254
11.1.1 模糊域的 EQUAL 关系.....	254
11.1.2 EQUAL 关系不同定义的实例.....	255
11.2 模糊函数依赖.....	256
11.2.1 定义.....	256
11.2.2 模糊函数依赖的实例 1.....	257
11.2.3 模糊函数依赖实例 2.....	259
11.3 模糊函数依赖的推导公理系统.....	261
11.3.1 公理系统.....	261
11.3.2 有效性.....	261
11.3.3 完备性.....	263
11.4 无损连接的分解.....	266
11.4.1 模糊无损连接的分解的定义.....	266
11.4.2 模糊无损连接分解的条件.....	268
11.4.3 模糊追赶算法.....	272
第十二章 动态函数依赖	276
12.1 关系的动态模型.....	276
12.2 动态函数依赖.....	277
12.2.1 动态约束与静态约束.....	277
12.2.2 动态约束的两个例子.....	277
12.2.3 更新与“作用关系”.....	281
12.2.4 动态函数依赖的形式化定义.....	282
12.2.5 二分的动态函数依赖.....	283
12.3 动态函数依赖模式的闭包.....	284
12.3.1 定义.....	284
12.3.2 动态函数依赖模式闭包的计算方法.....	285

12.3.3	两个动态函数依赖模式的等价	286
12.4	动态映射	286
12.4.1	四种动态映射	286
12.4.2	动态映射性质 1	287
12.4.3	动态映射性质 2	288
12.4.4	例子	289
12.4.5	动态映射性质 3	290
12.4.6	动态映射性质 4	291
12.4.7	例子	292
第十三章	关系的“老化”	293
13.1	稳定关系的年龄	293
13.1.1	稳定的关系序列	293
13.1.2	age-K 闭包	294
13.2	age-K 闭包的计算方法	294
13.2.1	作用属性、作用约束与作用关系	294
13.2.2	age-K 闭包计算方法(一)	296
13.2.3	二分的函数依赖	297
13.2.4	age-K 闭包的计算方法(二)	300
13.3	关系的“老化”	302
13.3.1	age-K 闭包序列的“收敛”	302
13.3.2	“成年”关系	303
13.3.3	对动态映射封闭的最小闭包	305
13.3.4	σ 公理	306
13.3.5	age-K 闭包序列“极限”推导算法证明	308
13.4	任意关系的年龄	312
13.4.1	任意关系序列中元组的年龄	312
13.4.2	关系的生命力	313
第十四章	目标投影视图的动态模式	316
14.1	关系中目标的体现	316
14.1.1	关系中的目标	316
14.1.2	目标属性集合的形式化定义	317
14.1.3	目标属性集合的识别	318
14.1.4	正则动态扩充	319
14.2	目标投影视图的约束	321
14.2.1	动态函数依赖族的投影	321
14.2.2	包含目标投影视图的最小动态函数依赖族	324
14.3	目标投影视图的更新	325
14.3.1	目标-投影-视图模式(O-P-V 模式)	325
14.3.2	可更新的 O-P-V 模式	326

14.3.3	O—P—V 模式可更新的充要条件	327
14.3.4	O—P—V 模式可更新的简易判定条件	330
14.3.5	正则动态扩充的目标投影视图的可更新性	331
14.3.6	可更新视图是函数依赖族的条件	334
第十五章	查询优化	335
15.1	查询优化概述	335
15.1.1	查询优化的目的	335
15.1.2	关系演算与关系代数的进一步的性质	336
15.2	查询优化的一般策略	344
15.2.1	语法树	344
15.2.2	关系代数的等价变换	345
15.2.3	关系代数表达式的优化算法	346
15.3	等式合取查询	348
15.3.1	合取查询的定义	348
15.3.2	等式合取查询的同态映射	349
15.3.3	等式合取查询的包含问题	351
15.3.4	自同态与极小化	352
15.4	稠密域不等式合取查询	354
15.4.1	不等式合取查询的特点	354
15.4.2	G(L)图	355
15.4.3	序等价赋值	356
15.4.4	稠密域不等式合取查询的包含问题	358
15.4.5	不等式合取查询中的同态条件	361
15.4.6	半开区间不等式合取查询	362
15.5	必要常数与极小化	364
15.5.1	必要常数及不必要常数	364
15.5.2	查询的“扰动”	365
15.5.3	“扰动”对等式的效果	367
15.5.4	等价查询与必要常数的关系	368
15.6	离散域不等式合取查询	368
15.6.1	保合式映射	368
15.6.2	基本定理	370
15.7	离散域不等式合取查询包含的算法	371
15.7.1	赋值的特征值	371
15.7.2	m 位 m 进制数是特征值的充要条件	372
15.7.3	单个域包含问题的判定算法	373
15.7.4	多个域时查询包含问题的判定算法	375

第一章 关系模型

1.1 基本定义

在数据库中存储着许许多多的数据,这些数据都是对客观事物的记载。作为一个数据库,它显然不能包罗万象地记载客观事物的一切性质,而只能记载我们在某个问题上所关心的那些性质,这样一些性质特称为“属性”。例如,对于一个学生,姓名、年龄、性别、班级等是他的“属性”。再如一本书,书名、作者、出版社、价格、内容摘要等也都是它的“属性”。

属性可用属性名直接表示,也可用 $A, B, C, \dots$ 等排在前面的大写英文字母来代表。每个属性都对应一个“域”,属性 A 的域记作 $dom(A)$ 。例如: $dom(\text{性别}) = \{\text{“男”, “女”}\}$, 在一个有关学生的数据库中 $dom(\text{年龄}) = \{20, 21, \dots, 35\}$, 而 $dom(\text{姓名}) = \{\alpha\beta\gamma | \beta, \gamma \text{ 是任意汉字或空格, } \alpha \text{ 是可作为姓的汉字}\}$ 等等。

任何域中的元素都称为“常数”,所有常数的集合记为 $\mathcal{C}$ 。

属性的有限集合称作“关系模式”,用 R, S 等来表示,例如 $R = \{\text{姓名, 年龄, 性别}\}$ 就是一个关系模式。

关系模式 R 上的一个元组 u 是 R 到 $\mathcal{C}$ 的一个映射,使得若 $A \in R$, 则 $u[A] \in dom(A)$ 。例如: $u[\text{姓名}] = \text{李江平}, u[\text{年龄}] = 27, u[\text{性别}] = \text{男}$, 则 $u = \langle \text{李江平}, 27, \text{男} \rangle$ 就是上述关系模式 R 上的一个元组。元组一般用 u, v, w 等小写英文字母来代表。

关系模式 R 上的所有元组的集合记作 $Tup(R)$ 。 $Tup(R)$ 的任何一个有限子集 r 都称作是 R 上的一个关系,或称 R 的一个实例。例如: $r = \{\langle \text{李江平}, 27, \text{男} \rangle, \langle \text{张铁诚}, 21, \text{女} \rangle, \langle \text{刘伟英}, 35, \text{男} \rangle\}$ 就是 R 上的一个关系。关系一般用小写英文字母 r, s 等来代表。

显然每个关系都可用一个矩形的表来表示,例如以上的关系 r 可以用表 1.1 所示的表来表示。其中第一行是属性名,从第二行起每一行是一个元组,而每一列则对应一个属性。

姓名	年龄	性别
李江平	27	男
张铁诚	21	女
刘伟英	35	男

表 1.1

关系模式 R 上的所有实例的集合记作 $Ins(R)$ 。设 $\mathcal{U} = \{A, B, C, \dots\}$ 是我们在数据库中关心的所有属性的集合,简称“全属性集合”,我们要求 $\mathcal{U}$ 是有限的,若 $R_1, R_2, \dots, R_n$ 都是 $\mathcal{U}$ 的子集,从而它们都是关系模式,且 $R_1 \cup R_2 \cup \dots \cup R_n = \mathcal{U}$, 则称 $\Omega = \{R_1, R_2, \dots, R_n\}$ 为 $\mathcal{U}$ 上的数据库模式。例如: $\mathcal{U} = \{\text{姓名, 年龄, 性别, 班级, 书名, 作者, 价格}\}, \Omega = \{R_1, R_2, R_3\}$, 其中:

$$R_1 = \{\text{姓名, 年龄, 性别, 班级}\}$$

$R_2 = \{\text{姓名, 书名}\}$
 $R_3 = \{\text{书名, 作者, 价格}\}$

就是 $\mathcal{U}$ 上的一个(表示借书情况的)数据库模式。 Ω 上的一个映射 ω , 使得每个 $R_i \in \Omega$, 都有 $\omega(R_i)$ 是 R_i 上的一个关系 r_i , 则称 ω 是 Ω 的一个实例, 也称 ω 是 Ω 上的一个“关系数据库”, 简称“数据库”, 也可记作 $\omega = \{r_1, \dots, r_n\}$ 。

例如: 对上述数据库模式 $\Omega = \{R_1, R_2, R_3\}$, 令 $\omega(R_1) = r_1$ (如表 1.2 所示), $\omega(R_2) = r_2$ (如表 1.3 所示), $\omega(R_3) = r_3$ (如表 1.4 所示), 则 $\omega = \{r_1, r_2, r_3\}$ 是一个关系数据库。

姓名	年龄	性别	班级
李江平	27	男	86-1
张铁诚	21	女	91-1
刘伟英	35	男	80-1

表 1.2

姓名	书名
李江平	人工智能
张铁诚	编译原理
张铁诚	人工智能

表 1.3

书名	作者	价格
人工智能	常胜	7.50
编译原理	陆杰	6.70
数据结构	赵英	10.20
形式语言	丁模	8.40

表 1.4

1.2 关系运算

我们使用一个关系数据库 $\omega = \{r_1, \dots, r_n\}$, 就要对它进行诸如存入、检索、更新等操作, 而这些操作是建立在关系运算基础上的。关系运算在传统上分为: 对元组的运算、关系代数及关系演算三种。

1.2.1 对元组的运算

1. 投影 $u[X]$

若 R 是一个关系模式, $X \subseteq R, u \in \text{Tuple}(R)$, 则 u 在 X 上的投影, 记作 $u[X]$, 是 X 上的一个元组 v , 满足对每个 $A \in X$, 都有 $v[A] = u[A]$ 。