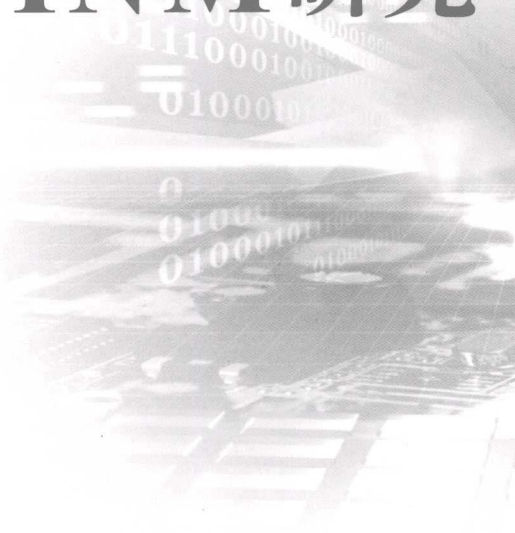
INM

INM

信息网模型INM研究

胡 婕 刘梦赤 著



科学出版社

信息网模型 INM 研究

胡 婕 刘梦赤 著

科学出版社
北 京

版权所有，侵权必究

举报电话：010-64030229；010-64034315；13501151303

内 容 简 介

本书围绕非结构数据元数据的语义建模过程中涉及的问题展开研究与讨论. 主要内容包括：非结构数据的相关概念及其在语义表示和搜索方面存在的问题和研究现状；针对存在的问题提出新的数据模型 INM，介绍 INM 的基本概念并将它与面向对象模型和角色模型进行比较，总结 INM 的特色；介绍 INM 的模式语言和实例语言的语法，重点研究其形式化语义；分析基于不同模型和针对不同逻辑结构数据的查询语言的特点及存在的问题，介绍专门针对 INM 所设计的查询语言 IQL，重点研究其语法和形式化语义，总结 IQL 的特色；介绍以 INM 为概念模型的数据库管理系统 INM-DBMS 原型的系统结构及设计与实现；最后以两个典型的领域全面地展示如何用 INM 建模及它们在 INM-DBMS 中的应用.

本书通过实例说明原理，对从事数据库、信息建模以及语义网研究的专业教师和科研人员具有重要的参考价值，还可以作为计算机、信息技术等专业的大学生、研究生学习、研究的参考资料.

图书在版编目(CIP)数据

信息网模型 INM 研究/胡婕，刘梦赤著. —北京：科学出版社，2011.6

ISBN 978-7-03-031203-7

Ⅰ.①信… Ⅱ.①胡…②刘… Ⅲ.①信息网络-数据模型-研究 Ⅳ.①G202

中国版本图书馆 CIP 数据核字(2011)第 100287 号

责任编辑：曾 莉/责任校对：董艳辉

责任印制：彭 超/封面设计：苏 波

科学出版社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

武汉市科利德印务有限公司印刷

科学出版社发行 各地新华书店经销

*

2011 年 6 月第 一 版 开本：B5(720×1000)

2011 年 6 月第一次印刷 印张：14 1/4

印数：1—2 000 字数：280 000

定价：45.00 元

(如有印装质量问题，我社负责调换)

前　言

随着互联网技术的迅猛发展，各行各业面临的信息呈现爆炸式增长，非结构化数据的管理被广泛认为是信息技术产业亟待解决的一个重要问题. 美林公司的统计资料表明，全球 15%左右的信息有效地存储在各种类型的结构化数据库中，但是还有 85%的信息是非结构化的. 如何有效地管理大量的非结构化数据，对其进行有效的分析、储存、管理和搜索，这些相关理论研究只是处于起步阶段. 其中，最突出的是非结构数据的搜索问题. 一方面现有的方法不能解决非结构数据的搜索问题，另一方面管理结构化数据的数据库技术又不能直接应用在非结构化数据上，因此迫切需要崭新的非结构数据管理的概念、方法、技术和理论从根本上解决这一问题.

本书以作者攻读博士学位期间所在实验室承担的国家杰出青年科学基金(外籍)项目——“非结构数据管理的理论基础和系统实现”为研究背景，主要围绕非结构数据的元数据语义建模这一目标展开讨论. 从数据模型、建模语言、查询语言、系统实现和应用多个方面深入系统地研究了非结构数据的理论模型、逻辑基础和方法论.

全书共分为 6 章. 第 1 章介绍非结构数据的相关概念，分析非结构数据在语义表示和搜索方面存在的问题，国内外主流数据模型的研究现状以及它们在表示非结构数据元数据方面存在的不足. 第 2 章首先以具体的应用示例讨论分析对象模型(OMs)和角色模型(RMs)存在的问题，提出新的数据模型——信息网模型(简称 INM)；然后介绍 INM 的主要概念及其如何解决其他模型存在的问题；最后对面向对象模型、角色模型和 INM 进行比较并总结了 INM 的特色. 第 3 章介绍 INM 的模式语言和实例语言的语法，针对建模语言简洁、高度集成但是能表达丰富的语义的特点，重点研究了其形式化语义. 第 4 章首先系统地研究基于不同模型和针对不同逻辑结构数据的主流查询语言，如 OQL、XPath、XQuery、GOQL、GraphQL 等，对它们的特点进行概括和总结并分析它们存在的问题及设计新的查询语言的必要性；然后介绍专门针对 INM 所设计的查询语言(简称 IQL)，阐述 IQL 的语法和形式化语义；最后对 IQL 的次要功能进行说明并总结 IQL 的特色. 第 5 章介绍以 INM 为概念模型所设计的数据库管理系统(简称 INM-DBMS)原型，它提供了完善的建模语言和查询语言对元数据进行定义、操纵、管理和查询. 研究 INM-DBMS 的系统结构及通信层、逻辑层和物理层各个功能模块的设计与

实现，介绍系统的开发和运行环境. 第 6 章以“DBLP”和“电影多媒体”两个典型的领域全面地展示如何用 INM 建模及它们在 INM-DBMS 中的应用，最后以具体的应用为例介绍原型系统中 Java 客户端的功能.

本书具有重要的应用价值. 书中以对非结构数据需求较大的领域（如教育机构、政府部门、科研机构、娱乐圈）相关的非结构数据（如科研项目、论文、学术会议、电影、音乐、奖项）为研究对象，分析现有非结构数据语义搜索存在问题和提出新的数据模型的必要性. 此外，对模型的建模语言和查询语言的设计始终坚持实用、直接、自然、与现实世界一一对应的原则. 这些特点使得本书提出的模型可以推而广之应用于很多类似的领域.

本书具有较为重要的理论价值. 书中的许多研究成果，特别是对现有语义数据模型表达能力和特点的分析和总结，对新模型及其建模语言和查询语言的逻辑基础和形式化语义的研究都是先期研究从未涉及的，具有原创性. 本书的研究涉及数据库、信息建模、语义网等多个学科领域，因此从这个角度讲，本书极大地丰富了这些学科领域对非结构数据管理的探索研究，对学科的深度发展将产生推动作用.

本书的出版得到国家杰出青年科学基金（外籍）项目（“非结构数据管理的理论基础和系统实现”）和 973 计划项目（“需求工程——对复杂系统的软件工程的基础研究”）的资助，也得到我的博士导师刘梦赤教授的悉心指导和鞭策、鼓励. 此外，书中参考了许多学者的研究成果，在此一并表示衷心的感谢.

限于作者学识水平，书中不足和疏漏之处在所难免，敬请同行和读者批评指正.

胡　婕

2011 年 2 月 23 日

此为试读，需要完整PDF请访问：www.ertongbook.com

目　录

第1章 绪 论

1.1 研究背景

本书的研究主要来源于国家杰出青年科学基金(外籍)项目——“非结构数据管理的理论基础和系统实现(No. 60688201)”.

非结构数据是指那些无法用计算机识别其结构的计算机化的信息. 这类数据通常包括非结构文本,字处理文本文件,PowerPoint 演示文件,图像、音频以及视频文件等;它主要出现在电子邮件、备忘录、备注、新闻、聊天记录、报告、信件、综述、白皮书、市场资料、研究论文、演示稿、网页中. 非结构数据广泛地存在于 PC、服务器、内联网以及互联网上.

根据美林公司估计全球 85%的信息是非结构化的,目前还没有成熟的技术和产品对其进行有效的存储、管理、分析和搜索,其相关理论研究也只是处于起步阶段. 该项目以 PC、服务器、内联网和互联网的非结构数据管理为需求导向,目的是要设计出一种新型的数据模型来描述非结构数据元数据之间的各种语义关系,基于该模型设计相应的数据定义、数据操纵、数据查询语言并研究其逻辑基础,以有效地存储、管理、维护和推理非结构化数据,并基于语境、语义和元数据之间的各种复杂关系进行搜索.

在项目的研究过程中,首先考虑非结构数据的搜索问题. 目前主要有三种方法:

(1) 关键字检索. 首先在文档中建立关键字的倒排索引,然后根据用户提供的关键字进行搜索,返回关键字所对应的网页或文档列表,最著名的是 google 和 baidu 等. 这种方法的主要问题是缺乏语境及语义支持,不考虑网页文档之间的各种关系,搜索的结果往往是海量但不相关的数据.

(2) 文档分类. 首先用树形层次分类目录对网页和文档进行分类,用户可以沿着分类目录的层次结构找下去,直到找到他们所需要的网页或文档;也可以直接用关键字搜索,返回关键字对应的网页或文档列表及其所属的分类;还可以在某个特定的分类下用关键字搜索,返回关键字在该目录下对应的网页或文档列表. 最著

名的是由 Yahoo 创建的 Web 目录[①]和开放目录计划(DMOZ)[②]等. 这种方法在一定程度上能够表示非结构数据的结构化语义,主要问题在于分类目录靠人工建立和维护,效率比较低下;分类目录的逻辑结构是树结构,信息的分类不符合自然的思维习惯,用户往往不能决定如何进行文档分类,而且在查找时所选择类别不一定正好是文档所属的类别,因而往往找不到需要的信息. 此外,分类目录所能表达的语义有限,因而无法提供丰富的语义搜索功能.

(3) 文档内容智能化. 用基于统计和机器学习方法的自动分类工具从非结构数据中自动选择重要的概念,生成元数据将文档进行自动分类,建立和维护这些分类层次,并提供复杂的用户接口来根据分类层次浏览文档. 其主要问题是无用户参与的这些分类系统的精确率很低,而且无法获取非结构数据元数据之间各种自然复杂的联系.

非结构数据的语义搜索是在获得了被搜索数据元数据的语义基础上,通过对语义进行直接、自然地表示和处理以后,才能使得搜索的结果在意义上,而不仅仅是在语法或结构上满足搜索的需求. 用户经常会变换着使用各种关键字的组合,希望能从搜索引擎那里得到所需要的结果,可往往是无功而返. 例如,当一个不知道武汉大学校长是谁的用户想找有关武汉大学校长的信息时,输入"武汉大学校长"得到的只是一些匹配"武汉大学校长"关键字的网页或文档. 搜索引擎并不能直接告诉用户谁是武汉大学校长,用户需要自己在结果中做"第二次"、"第三次",甚至"第 n 次"的过滤才能满足自己的需求. 用户希望系统可以直接把"顾海良"告诉用户,虽然"顾海良"在关键字意义上是不匹配的. 这里的关键在于,搜索引擎必须能够理解"武汉大学校长"是非结构数据元数据中的一个概念,而"顾海良"是这个概念的实例. 也就是说,系统必须能够从语义上处理查询. 因此,如何自然、直接地表示非结构数据元数据的语义是需要解决的首要问题.

通过对教育机构、政府部门、科研机构、科研项目、论文、学术会议、电影、音乐和奖项等领域非结构数据的元数据进行调查、研究和建模,会发现元数据之间存在各种复杂、动态的关系,而且通过这些关系可以自动获得元数据的上下文语境信息. 利用元数据之间的这些关系及语境信息可以得到更精准、更有意义的搜索结果.

以与大学领域相关的元数据为例,"大学"、"校长"、"教授"、"博士生"、"科研项目"这五个概念之间的语义可以抽象为:大学有校长、教授、博士生三类人;校长是人的职务,教授是在大学工作的一类人的职称,博士生是就读于某所大学的学生的身份;校长、教授、博士生还分别是拥有校长职务、教授职称、博士生身份的一类人;

① Yahoo Directory. Http://dir.yahoo.com/.

② Domz. Http://www.dmoz.org/.

教授指导博士生并主持科研项目;反之,博士生的导师是教授;科研项目有主持人和参与者. 对于某个具体的大学而言,元数据对象之间的语义可以抽象为:武汉大学的校长是顾海良,他同时也是武汉大学的教授,武汉大学还有博士生张新和郑吉伟;反之,校长顾海良的职务是武汉大学的校长,作为校长他的任职开始年份是2008年;教授顾海良在武汉大学工作其职称是教授,作为教授他指导博士生张新并且主持了项目"斯大林社会主义建设理论与实践研究";博士生张新就读于武汉大学,其身份是博士生,作为博士生他的导师是顾海良并且参与了项目"斯大林社会主义建设理论与实践研究";国家"九五"社科基金资助项目"斯大林社会主义建设理论与实践研究"的主持人是顾海良,参与者有张新和郑吉伟. 如果上述元数据的语义已经明确地表示如下:

```
大学 武汉大学[
        校长:顾海良,
        教授:顾海良,
        博士生:{张新,郑吉伟}]
{校长,教授} 顾海良[
        职务:武汉大学.校长[开始年份:2008],
        工作单位:武汉大学[职称:教授[
            指导博士生:张新,
            主持项目:斯大林社会主义建设理论与实践研究]]]
博士生 张新[
        就读于:武汉大学[身份:博士生[
            导师:顾海良,
            参与项目:斯大林社会主义建设理论与实践研究]]]
国家"九五"社科基金资助项目 斯大林社会主义建设理论与实践研究[
        主持人:顾海良,
        参与者:{张新,郑吉伟}]
```

那么要搜索"武汉大学的校长"、"顾海良的职务"、"顾海良的博士生参与的项目的主持人和参与者"等就能非常直接、自然地表示如下:

```
武汉大学//校长:$x
顾海良//职务:$x
顾海良//指导博士生:$x//参与项目:$y[主持人:$z,参考者:$t]
```

搜索的结果会更有意义。

以上述需求为导向,下面将深入系统地研究国内外主流数据模型,希望能直接用它们建模. 考虑上述例子,"校长"这一概念除了表示一类人如顾海良,还表示从"武汉大学"到"顾海良"之间的一种关系,同时还表示顾海良在武汉大学所扮演的角色. 这三者之间存在内在的联系,通过这些内在的联系可以获得元数据的上下文,进而可以自然地表示上下文语境信息. 但是发现现有主流数据模型都无法同

时直接、自然地支持元数据的复杂动态关系、多刻面、动态演化和上下文语境信息表示和访问这几个非常重要的特性. 所以,研究出一种新的能直接、自然地表达非结构数据元数据之间的各种语义关系的数据模型,从而基于该模型设计出其相应的建模语言和查询语言,研究它们的逻辑基础,并基于语义、语境和非结构化数据元数据之间的各种复杂关系进行搜索成为本书的研究重点.

1.2 国内外研究现状

数据模型的研究并不是一个新课题,从 20 世纪 70 年代开始它就成为数据库领域的热点研究问题之一.

迄今为止最著名、使用最广泛的数据模型是 1970 年由 Codd 提出的关系模型(Relational Model,简称 RM)[1-5]. 关系模型的主要优点在于它建立在严格的数学理论基础上,概念单一,无论实体或实体之间的联系都用二维表格结构表示. 但是"概念单一"也使得关系模型无法直接、自然地模拟现实世界中实体之间的各种复杂联系,更加无法支持上下文语境信息的表示. 同时,实体联系的语义关系无法在模式中显式地表示出来,而要求用户通过自己的理解从语义上操纵这些联系. 这对于建模人员的要求比较高,客观上需要有更高级的概念语义模型来辅助建模.

为了满足概念建模的需求,1975 年 Peter P. Chen 提出了实体-联系模型(Entity Relationship Model,简称 ER)[6-15],ER 模型将现实世界中的概念抽象成实体、联系和属性,概念之间的所有语义关系都用这三者表达. 它简化了信息建模的过程,在数据库概念层设计方面得到了广泛的应用,也是迄今为止在数据库领域使用最广泛的语义模型之一. 从 1979 年开始很多研究者在各个方面对 ER 模型进行了扩展,并提出了 EER 模型[16-32]. 例如,文献[18,33,34]等对抽象机制如泛化(generalization)和特化(specialization)进行了扩展;文献[35]扩展了三元关系和复合属性;文献[36,37]等在基数约束(cardinality)和一般完整性约束方面进行了扩展. 但是无论是 ER 模型还是 EER 模型,对于"联系"的表示,最复杂只能支持带属性的联系. 而对于复杂联系,如层次结构的联系和联系的派生类等都无法表示.

除了 ER 模型外,同时期有些学者还提出了其他语义模型. 1976 年 Kerschberg 等提出了 Functional Data Model,简称 FDM[38],FDM 不支持聚合和分组,对象之间的联系直接用属性表示,它是第一个基于属性的语义模型,FDM 的最大特点是直接参照函数操纵数据属性. 后来有些研究者对 FDM 进行了扩展,例如,Shipman 在 FDM 中引进了派生模式的概念[39];Dayal 等提出了 FDM 模式基于图的非形式化表示[40]. 1978 年 Hammer 等提出了 Semantic Data Model,简称 SDM[41,42],SDM 提供了丰富的派生属性和派生子类基本元素,它是第一个支持分组构造函数和派生模式的语义模型. SDM 比 ER 模型更复杂、表达能力也更强大.

此为试读,需要完整PDF请访问: www.ertongbook.com

FDM 和 SDM 与 ER 模型的最大不同在于，ER 模型是面向联系而 FDM 和 SDM 是面向属性的，它们简化了 ER 中的联系，因此同样也无法支持复杂联系.

随着面向对象的程序设计的流行，20 世纪 80 年代开始，面向对象模型(Object-Oriented Model，简称 OOMs)[43-56]成为研究的热点. 在各种面向对象模型中，将 ER 模型中的实体抽象为“对象”，并且主要围绕对象开展研究，如对象标识、复杂对象、对象分类、对象聚合、类泛化和特化、非单调继承、覆盖和重载等. 在传统的面向对象模型中，一个对象只能是一个类的实例[45,55]，这导致对象无法改变其类从属关系，从而只能表达对象的静态特性而无法表现其动态演化特性. 为解决这些问题，有些面向对象模型支持不相交类的多继承，但是多继承会导致子类的组合爆炸[57-59]. 为避免多继承子类的组合爆炸问题，有些面向对象的模型支持多分类，但是它们无法支持对象的上下文语境信息表示. 90 年代开始，各种针对传统的面向对象模型存在的这些问题所提出的数据模型成为研究的热点.

为了刻画现实世界中对象的动态演化、多刻面、上下文语境信息特性，研究者们提出了各种角色模型(Role Models，简称 RMs)[57-71]. 角色模型的最大特点是将面向对象模型中的类分为两种：对象类和角色类. 对象类用来表示对象的静态特性；角色类用来表示对象类所扮演的角色，一个对象可以扮演多个角色，角色主要强调对象的动态、多刻面特性. 角色类和对象类都可以有分类层次，对象类的分类层次表示对象的静态分类，而角色类的分类层次表示对象的动态分类. 对象类的继承体现在模式级别，而角色类的继承体现在实例级别. 此外，角色类还支持简单的上下文语境信息访问. 角色模型主要存在以下两个方面的问题：

(1) 为了能够表示对象的多刻面和上下文语境信息特性，现实世界中一个实体的信息只能分散地表示在一个对象实例和多个层次结构的角色实例中，这种表示和现实世界的概念并不完全对应. 也就是说，尽管它能表现对象的动态演化、多刻面、简单的上下文语境信息访问特性，但是表示并不自然.

(2) 它们只能孤立地表示对象所扮演的角色而无法表示对象在关系所处的语境中所扮演的角色.

Terry Halpin 等在面向对象模型的基础上提出了对象角色模型(Object Role Model，简称 ORM)[72-81]，ORM 将数据描述为不可再分割的事实，它是一种基于事实的模型. ORM 主要有两个特点：

(1) 它不显示地表示属性，属性和关系都统一地表示为角色，将现实世界的概念抽象为具有角色的一组对象.

(2) ORM 用直观的图形或者自然语言来分析对象和角色之间的语义，对用户来说使用起来更方便.

需要注意的是对象角色模型中的“角色”和角色模型中“角色”是两种不同的概念，对象角色模型中的“角色”是属性和关系的统一表示，而角色模型中“角色”通常

是对象类的动态子类. 从建模的表现方式角度看,ORM 比 ER 模型和 OOMs 更直观. 但是由于 ORM 设计者的初衷是针对商业领域数据库系统的概念,为数据建模提供更简单、直接、灵活的方式,因此在表达复杂动态联系、上下文语义语境相关表示等方面与上述模型相比并没有优势.

1995 年,彭智勇等提出了对象代理模型(Object Deputy Model)[82-85],对象代理模型通过引入代理类和代理对象的概念对传统的面向对象数据模型进行了扩展. 代理类用来描述代理对象的模式,代理对象用来扩展和定制其源对象. 对象代理模型能够提供特化、泛化、聚合和分组等抽象机制,实现对象视图、角色多样性及对象迁移等功能,因此它比传统的面向对象模型更加灵活. 但是用对象视图的机制将一个对象的信息拆分成多个代理对象来表达对象的多刻面和迁移特性与现实世界中的对象概念也不能一一对应,因此这种建模方法的主要问题是表示不自然.

综上所述,以上模型都无法同时直接、自然地支持复杂关系、多刻面、动态演化及上下文语境信息表示和访问这几个非常重要的特性.

1.3 主要研究内容

虽然语义数据模型的研究在数据库领域并不是一个新兴的研究方向,但是并没有专门针对非结构数据的元数据进行建模并且能够满足项目需求的数据模型能直接为我们所用. 本书主要围绕非结构数据的元数据建模这一课题开展了全面、深入、系统的研究,所涉及的内容主要包括数据模型、建模语言和查询语言及其逻辑基础、模型的实现及其应用. 主要研究成果包括以下内容:

(1) **数据模型**. 以非结构数据的元数据语义建模为需求导向,全面细致地分析了现有数据模型存在的问题,研究如何更自然、直接地表示元数据之间的各种复杂的关系和上下文语境信息. 提出了信息网模型(Information Networking Model,简称 INM)[86-89],它克服了现有数据模型存在的问题,能更好地满足项目的需求.

(2) **建模语言**. 针对 INM 的特点提出了相应的建模语言[89,90],该语言具有语法简单、集成度高、语义丰富、表达力强等特点. 此外,还研究了 INM 建模语言的逻辑基础,并对建模语言的语义进行了详尽的形式化描述.

(3) **查询语言**. 全面地分析了现有经典流行的查询语言和基于图结构数据模型的查询语言的特点,针对 INM 提出了一种新颖、功能强大、表达能力强的查询语言 IQL[91-93],详尽介绍了其语法和形式化语义.

(4) **模型的实现**. 介绍了项目组开发的基于 INM 的数据库管理系统(INM-DBMS)的设计与实现,包括其系统结构、系统设计和系统实现.

(5) **模型的应用**. 用 INM 对 DBLP 和电影多媒体建模并展示了它们在 INM-DBMS 中的应用.

第2章 信息网模型INM

本章以具体的应用示例分析了面向对象模型和角色模型存在的问题，并提出了一种新的数据模型——信息网模型，介绍了信息网模型的相关概念并总结了它与其他模型相比较的优势及其创新点.

2.1 问题分析

经过全面、深入地调查研究大学相关领域非结构数据的元数据对象之间的语义关系，可以发现元数据对象之间存在着各种复杂的关系，基于这些关系对象扮演着不同的角色，通过这些关系中的角色可以自动地获得元数据对象的上下文，进而能自然地展现各种上下文语境信息. 1.2 节所述的数据模型，如关系模型、ER 模型、EER 模型、FDM、SDM、面向对象模型、角色模型、对象角色模型、对象代理模型，都过度简化甚至忽略了这些关系，或者只考虑对象所扮演的角色及角色的属性而不考虑角色所在的关系. 因此，它们无法直接、自然地模拟对象与对象、对象与关系以及关系与关系之间的各种关系，也无法支持上下文语境信息的表示和访问. 从这几个方面来说，它们只能对现实世界进行部分建模而无法在现实世界和数据模型之间建立一一对应的关系.

下面以一个具体的例子来分析面向对象模型和角色模型在上述几个方面存在的问题. 选定这两个模型，主要原因是本书所提出的模型是基于面向对象模型的，同时还借鉴了角色模型中的一些思想. 换句话说，面向对象模型和角色模型与本书所提出的模型最接近.

考虑大学元数据对象建模的应用示例：大学有各种各样的人，如校领导、教师、学生. 校领导有任期、办公室、开始年份并且可特化为校长和副校长；反之，若某个人扮演某个大学的校领导、校长或者副校长角色，他就在该校担任相应的职务. 教师有开始年份并且可特化为讲师和教授等；反之，若某个人扮演某个大学的讲师或者教授角色，他就在该大学工作并且职业是教师且获得相应的职称. 学生有学号并且可特化为研究生和本科生，研究生的导师是教授并且可以特化为硕士生和博

士生;反之,教授指导研究生. 若某个人扮演某个大学的学生、研究生、硕士生或者博士生角色,他就在该校学习并且获得相应的身份. 大学还可能有各种校队,校队有运动员和副教练. 运动员有开始年份并且他们必须是本科生;反之,若某个人扮演某个校队的运动员角色,他就是该校队的成员之一. 课程有学分、先行课、其授课者是教师、选课者是学生,并且可以特化为研究生课程和本科生课程;反之,课程有后续课程、教师讲授课程、学生选修课程. 本科生课程的选课者是本科生,研究生课程的选课者是研究生;反之,本科生选修本科生课程,研究生选修研究生课程.

面向对象模型主要强调对象分类、复杂对象、类泛化和特化、类层次和继承等. 有些面向对象模型要求一个对象只能是最相关类的实例,这导致对象无法改变其类从属关系,故只能表达对象的静态特性. 有些面向对象模型支持不相交类的多继承,它们虽然能改变类从属关系,但是多继承会导致子类的组合爆炸. 有些面向对象的模型支持多分类,它们能解决多继承所导致的子类组合爆炸,但是无法支持对象的上下文语境信息访问.

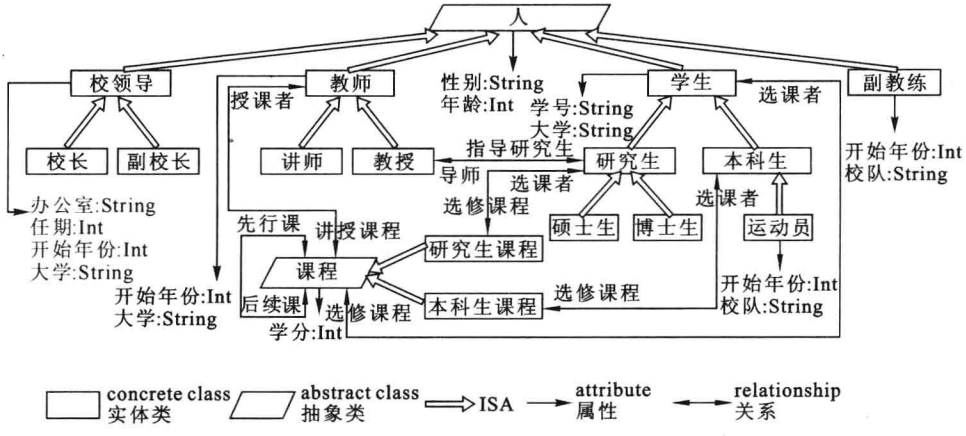
用面向对象模型对上述应用建模,可以用如图 2.1 所示来表示模式和实例. 如果想表示 Bob 既是 MIT 大学的校长和教授,又是女子篮球队副教练,可以用多分类来表示.

对象 Bob 的表示如图 2.1(b)所示,Bob 在面向对象模型中的表示如下:

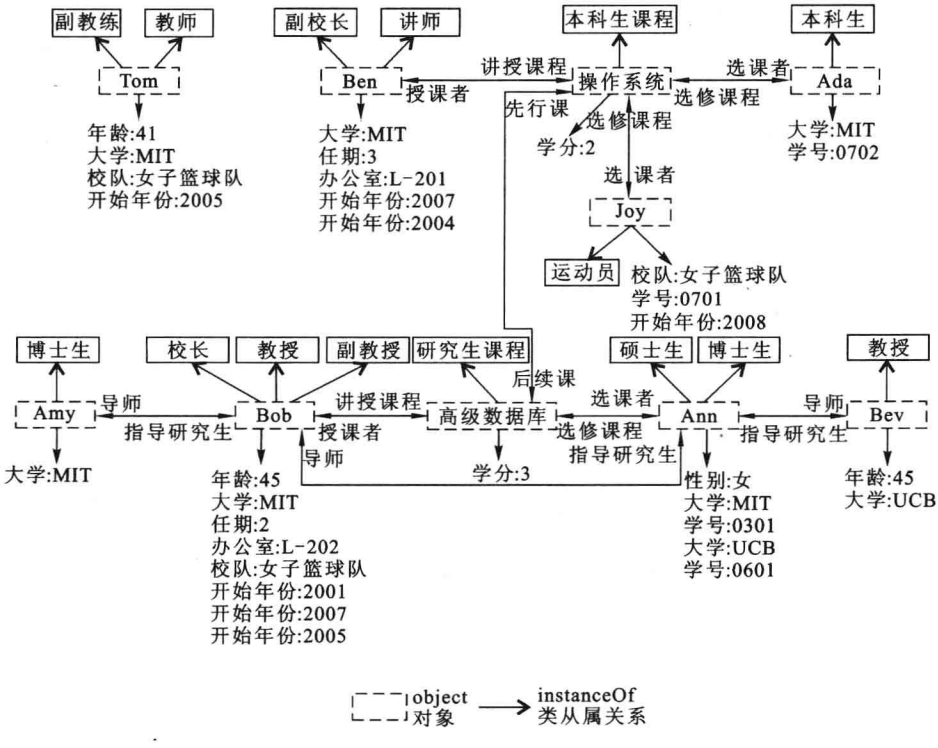
```
{校长,教授,副教练} Bob[
        年龄:45,
        大学:MIT,
        任期:2,
        办公室:L-202,
        校队:女子篮球队,
        开始年份:2001,
        开始年份:2007,
        开始年份:2005,
        讲授课程:高级数据库,
        指导研究生:{Ann,Amy}]
```

Bob 有属性“年龄”、“大学”、“任期”、“办公室”和“校队”,此外他还与“高级数据库”有“讲授课程”关系,与 Ann 和 Amy 有“指导研究生”关系. 注意:这里有三个同名的属性“开始年份”无法区分,即无法得知三个“开始年份”分别对应于“校长”、“教授”、“副教练”三个类中的哪一个. 对于 Ben,情况也类似.

再考虑另外一种情况,如果想表示 Ann 既是 MIT 大学的硕士又是 UCB 大学的博士,同样可以用多分类来表示. 对象 Ann 的表示如图 2.1(b)所示,Ann 在面



(a) 模式



(b) 实例

图 2.1 用面向对象模型建模示例

向对象模型中的表示如下：

```
{硕士生,博士生}Ann[
    性别:女,
    大学:MIT,
    学号:0301,
    导师:Bob,
    大学:UCB,
    学号:0601,
    导师:Bev,
    选修课程:高级数据库]
```

这里同样存在两个“大学”、两个“学号”和两个“导师”无法区分的问题，即无法确定 Ann 是哪所大学的硕士生及其对应的学号和导师，也无法确定她是哪所大学的博士生及其对应的学号和导师.

在角色模型中，根据对象演化迁移特性不同将子类分为两种：动态子类和静态子类. 静态子类的实例不会发生迁移，而动态子类的实例可能发生迁移. 对于上述应用示例，若将“研究生课程”视为“课程”的静态子类，那么一门课程如果不是研究生课程就不会演变成研究生课程；若将“学生”视为“人”的动态子类，那么一个人如果不是学生就可能会演变成学生. 在这两种情况中，子类的实例也是父类的实例，即“研究生课程”和“学生”的实例分别是其父类“课程”和“人”的实例. 动态子类在角色模型中也称为角色子类，它们可以形成角色类层次. 一个对象实例可以对应一个或多个角色实例作为其扮演的角色，并且每一个角色实例都可以有属性和关系. 对象扮演的角色可以视为这些属性和关系的上下文，所以角色模型支持简单的上下文语境信息访问.

用角色模型对上述应用示例建模，可以用如图 2.2 所示来表示模式和实例. 其中，“校领导”、“教师”、“学生”是“人”的直接角色子类，它们分别形成类层次结构：校领导→{校长，副校长}，教师→{教授，讲师}，学生→{研究生→{硕士生，博士生}，本科生}.

在图 2.2(b)中，Bob 是人的实例并且有五个角色实例表示其扮演的角色，它们分别是：副教练 Bob、校领导 Bob、校长 Bob、教师 Bob、教授 Bob. Bob 在角色模型中的表示如下：

```
人   Bob[年龄:45]
校领导   校领导Bob roleOf Bob
校长   校长Bob roleOf 校领导 Bob[
    大学:MIT,
    任期:2,
```


此为试读，需要完整PDF请访问：www.ertongbook.com