



“十一五”国家重点图书

数字时代图书馆学情报学研究论丛

本体与数字图书馆

Ontology and Digital Library

董 慧 著



WUHAN UNIVERSITY PRESS

武汉大学出版社



数字时代图书馆情报学研究论丛

“十一五”国家重点图书

国家自然科学基金委员会资助项目

“基于本体的数字图书馆检索模型” 成果之一

“基于数字图书馆的本体演化与知识管理研究”

(批准号: 70373047、70773087)

本体与数字图书馆

Ontology and Digital Library

董 慧 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

本体与数字图书馆/董慧著. —武汉:武汉大学出版社,2008.8
“十一五”国家重点图书
数字时代图书馆学情报学研究论丛
ISBN 978-7-307-06392-1

I. 本… II. 董… III. 数字图书馆—研究 IV. G250.76

中国版本图书馆 CIP 数据核字(2008)第 093156 号

责任编辑:杨 华

责任校对:刘 欣

版式设计:詹锦玲

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.com.cn)

印刷:武汉中远印务有限公司

开本:720×980 1/16 印张:22.125 字数:317千字 插页:4

版次:2008年8月第1版 2008年8月第1次印刷

ISBN 978-7-307-06392-1/G·1211 定价:38.00元

版权所有,不得翻印;凡购买我社的图书,如有缺页、倒页、脱页等质量问题,请与当地图书销售部门联系调换。

《数字时代图书馆学情报学研究论丛》

编 委 会

顾问: 彭斐章(武汉大学资深教授,信息管理学院博士研究生导师)
孟广均(中国科学院国家科学图书馆教授,博士研究生导师)
吴慰慈(北京大学资深教授,信息管理系博士研究生导师)
胡述兆(台湾大学图书资讯学研究所教授,博士研究生导师)
梁战平(中国科技信息研究所教授,博士研究生导师)
倪 波(南京大学教授,博士研究生导师)
黄长著(中国社会科学院研究员,学部委员)
冯惠玲(中国人民大学副校长,信息资源管理学院博士研究生导师)

Raymond von Dran (Professor and Dean, School of Information
Studies, Syracuse University)

Harry Bruce (Professor and Dean, Information School, University
of Washington)

主编: 陈传夫 马费成 胡昌平

编委: (按姓氏笔画排序)

马费成	方 卿	王新才	刘 荣	刘家真
朱玉媛	朱静雯	沈祥兴	肖希明	李 纲
吴 平	陈传夫	张玉峰	张李义	邱均平
何绍华	周 宁	罗紫初	胡昌平	查先进
曹 之	黄如花	黄先蓉	黄凯卿	董有明
董 慧	焦玉英	詹德优		

内 容 简 介

本书是国家 985 二期项目武汉大学信息资源研究创新基地研究项目和国家自然科学基金资助项目的成果之一,同时,也是作者多年来研究数字图书馆和本体论及其应用的结晶。本书从数字图书馆的定义、特点、作用和机制的范式演化分析了数字图书馆的内涵和本质;从元数据管理与应用、数字化信息组织结构、数字对象的标识和服务机制等,揭示了实现数字图书馆工程中应该应用的主要技术;从本体的定义、语义网的结构、本体的描述语言和本体的逻辑基础分析了本体的定义、各领域的应用定义和它的理论基础与基本实现,以及应用前景;从本体库的建设和维护、本体的检索策略、推理机制和处理引擎,分析了本体应用所使用的关键技术;作者通过多年的实践,针对应用本体技术解决问题时遇到的困难,提出了解决困难的途径和方法,即本体分子理论,对本体分子理论进行了详细的解析和应用实践,证明该理论具有可行性、可操作性和科学性;本书以基于本体数字图书馆检索模型为例,揭示了本体应用实现的五大关键技术细节和步骤。

本书结构严谨、内容新颖、布局合理、重点突出、实例丰富,是国内第一部将本体应用于数字图书馆建设中的专著,值得一读。

本书可作为大学和研究院相关专业高年级学生或研究生的教材,也可作为关注这一领域的专业人员的参考书。

总 序

“图书馆学情报学”是我国的习惯用法，是涵盖图书馆学、情报学、档案学、出版发行学等学科的名称。在我国台湾被称为“图书馆与资讯科学”，英文为 Library and Information Science。美国也用 Library and Information Studies 来称谓这一学科。

1807 年，德国学者马丁·施莱廷格（Martin Schrettinger, 1772 ~ 1851）首次使用了“图书馆学”这一概念，1808 年他又在《试用图书馆学教科书大全》中建立了以图书馆整理为核心的学科体系，标志着图书馆学学科正式诞生。

自 1887 年美国学者杜威（Melvil Dewey, 1851 ~ 1931 年）在哥伦比亚大学创办世界第一所图书馆学校，1930 年在卡内基基金的资助下芝加哥大学设立第一所图书馆学博士班课程以来，图书馆学开始走进大学殿堂，成为高等教育中的一个专业。

图书馆学教育在美国的兴起带动了全球图书馆教育的发展。1919 年英国在伦敦大学建立了图书馆学院。目前，美国有 56 所美国图书馆学会（ALA）认可的图书馆学院，每年招收图书馆与情报学学生 26 000 人左右。

在施莱廷格后的两个世纪，图书馆学科不断变化。特别是在 20 世纪 50 年代以来的冷战期间，美苏军备竞赛，两大阵营形成。苏联卫星上天，美国实施阿波罗计划，科技文献激增。科学家对文献信息的获取变得困难。一门新型学科——情报学应运而生。1963 年美国文献工作学会正式更名为美国情报学会（ASIS）。大量增设图书馆学与情报学硕士点、博士点。图书馆学课程表中也增加了大

量的情报学课程。

20 世纪 70 年代, 计算机技术在图书馆与信息工作中广泛应用, 自动化、地区性图书馆网络形成, 机读目录广泛应用, 国际图联将世界书目控制列为核心计划。图书馆学 (Library Science) 发展为“图书馆与情报学” (Library and Information Science), 后来又进一步演变为“图书馆与情报研究” (Library and Information Studies)。

20 世纪 80 年代高新技术迅速发展, 信息时代到来。美国里根政府实施星球大战计划, 欧洲实施尤里卡计划等。联机图书馆系统广泛建立, 并扩展至世界主要发达国家。商业性联机数据库如 ORBIT, DIALOG 发展迅速, 图书馆与情报职业面临挑战。为适应信息时代要求, 国际上图书馆学情报学专业开始调整。国际上有较多大学将图书馆学院易名为图书馆与情报学院或信息研究学院, 图书馆学、情报学在硕士、博士层次合二为一。

20 世纪 90 年代, 全球进入后信息时代——数字时代到来。克林顿政府开始实施国家信息基础设施计划 (NII)、全球信息基础设施计划 (GII)。新一代互联网投入使用。欧美初步建成信息社会, 全球进入无缝信息环境。世贸组织建立和一揽子贸易协定生效, 使全球经济一体化并逐步进入知识经济时代。各国继续加强图书馆学、情报学学科调整。图书馆学、情报学学科内容向情报科学汇集。

进入 21 世纪以来, 国际上信息管理学科变化很快。自雪城 (SYRACUSE) 大学将学院更名为信息研究学院 (The School of Information Studies) 后, 在美国立即出现了 iSchool 的浪潮。伊利诺依斯大学、华盛顿大学、密歇根大学、匹兹堡大学、加州大学伯克利分校、北卡罗来纳大学等知名大学的图书馆与情报学院宣称自己为 iSchool。这些 iSchools 通过宪章组成 I-Schools 联盟 (ISG)。目前共有 20 所美国的大学加入联盟 (联盟宪章不允许超过 25 个)。iSchool 强调信息、技术与人的关系 (relationship between information, technology, and people)。iSchool 的标准包括: 必须有杰出的研究和杰出的博士教育; 必须能在科学、企业、教育与文化进步过

程中提供任何形式的信息所需的专门技术；必须能提供信息技术及其应用、信息使用与用户方面的专门知识。2004—2006 年的联盟领导委员会协调人是雪城大学信息研究学院的 Raymond von Dran 院长，2006—2007 年将由匹兹堡大学信息学院院长 Ron Larsen 担任。联盟成员的标准主要强调研究即实质性承担研究活动（三年中每年研究支出达到 100 万美元），同时，致力于培养未来的研究者（通常通过研究型的博士点），引领推动信息职业领域。

国际上图书馆与情报学科的发展表现出明显的特征：研究范围由传统的图书馆领域扩大到信息领域（information field），研究视野由实体的图书情报机构扩大到虚拟空间，研究对象由图书文献转向了信息内容。一系列相关学科如图书馆学、情报学、档案学、出版科学、信息管理与系统乃至数字商务汇集于信息科学（Information Sciences）下，从 F 而使图书馆学情报学研究发生了根本的变化。

武汉大学图书馆学科起源于 1920 年美国学者韦棣华女士创办的武昌文华大学图书科，档案专业起源于 1940 年的文华图书馆学专科学校的档案管理科。1978 年武汉大学创办科技情报学专业，后改为情报学专业。1983 年创办图书发行学专业，2002 年创办电子商务专业。1984 年经教育部批准建立武汉大学图书情报学院。2001 年更名为信息管理学院。图书馆学和情报学两个二级学科被国务院学位委员会批准为国家重点学科。“图书馆、情报与档案管理”被国务院学位委员会批准为一级学科博士学位授权点。教育部批准“武汉大学信息资源研究中心”为国家人文社会科学重点研究基地。信息产业部批准成立“国家信息资源管理（武汉）研究基地”。新闻出版总署批准建立“新闻出版总署武汉大学高级出版人才培养基地”。“网络信息资源开发与数字图书馆建设”被国家计委、教育部等批准为“十五”211 重点学科建设项目。建立一级学科博士后流动站。武汉大学信息资源研究创新基地被列为国家“985 二期工程”建设项目。一批院内校级重点研究基地如武汉大学四库学研究所、武汉大学中国科技评价中心、武汉大学政府信息

研究中心、武汉大学数字图书馆研究所、武汉大学出版发行学研究所、武汉大学图书馆情报学国际合作研究中心也在科研和人才培养中发挥着重要平台作用。

强调一级学科内学科群建设和学科协调发展是武汉大学图书馆与情报学科建设的基本目标。以图书馆学、情报学两个国家重点学科为龙头促进图书馆学、情报学、档案学、信息资源管理、出版发行学等学科的协调发展。

我们深刻认识到信息资源与自然资源、人力资源共同构成支撑现代经济社会发展的资源体系。信息资源是知识经济时代重要的国家战略资源，是实现经济和社会全面、可持续发展的基础条件。对信息资源的拥有、开发和利用水平，是衡量一个国家综合国力和国际竞争力的重要标志之一。消弭信息鸿沟、实现信息公平，是消除贫困、促进经济发展、构建和谐社会的重要条件之一。

信息资源管理人才培养是学院的基本任务。学院每年为国家培养本科生 260 名，硕士研究生 150 名，博士研究生 55 名左右。学院有一支知识结构和年龄结构合理的优秀学术队伍。这支队伍中有武汉大学人文社会科学资深教授 1 人，博士研究生导师 26 人，国务院政府特殊津贴专家 6 人，教育部新世纪优秀人才支持计划 3 人，武汉大学珞珈特聘教授 2 人。作为实现研究型学院建设目标的一部分，在教学的同时，广大教师承担了大量的科学研究任务。为了推动本学科领域的前进，分享他们的见解，在武汉大学出版社的大力支持下，并报有关部门批准，我们拟出版《数字时代图书馆情报学研究论丛》（简称《论丛》）。

为了编辑这套丛书，武汉大学邀请了国内外知名学者担任《论丛》的学术顾问，组建了主要由信息管理学院的博士研究生导师担任委员的编辑委员会。

《论丛》拟用 4 年时间出版著作共 20 卷。20 卷著作将分为三个系列：（1）学科年度进展。主要约请信息管理学院图书馆学系、档案与电子政务学系、信息管理科学系、现代出版系、信息系统与

电子商务系的有关教师和校外专家共同编写本学科的年度研究进展,主要有《图书馆学研究进展》、《情报学研究进展》、《档案学研究进展》、《出版学研究进展》、《信息资源管理学研究进展》;(2)个人学术专著。涉及图书馆、情报与档案管理基本理论研究、信息组织与检索、信息资源管理、信息资源建设与信息服务、文献编纂与出版、数字图书馆与信息系统工程等研究方向;(3)研究报告系列。我院研究人员共承担教育部哲学社会科学研究重大攻关项目、国家社会科学基金重点项目、教育部人文社会科学重点研究基地重大招标项目、国家自然科学基金项目、国家社会科学基金项目多项。特别是211项目和985项目,围绕数字信息资源开发与管理、数字信息资源服务与保障、信息资源公共获取与知识产权协调管理、数字图书馆关键技术与系统、资源与服务整合、信息构建与知识管理等主题正在进行探索。在信息构建的理论与方法、信息系统与资源整合、元数据知识表达、网络计量与参考、信息服务集成机制、信息资源与服务集成技术、媒体及数字出版、数字内容分销、信息资源的长期保存、商务信息流等关键领域力图实现图书馆学科在数字图书馆领域、情报学科在数字资源管理领域、档案学在数字化政务信息管理领域、出版发行学在数字出版与数字化分销、信息系统科学在集成系统以及数字化商务信息流研究方面取得研究成果。本系列将对部分研究结果进行报告。

丛书的出版是学院广大教师和研究人员辛勤探索的结果,在此,谨向严谨治学、辛勤耕耘的各位著作者表示感谢!对武汉大学出版社的支持表示感谢,对责任编辑严红女士在策划编辑过程中付出的艰辛劳动表示感谢。同时,还望广大读者不吝批评指正,共同推动图书馆学、情报学、档案学、出版发行学和信息资源管理学科的进步!

武汉大学信息管理学院院长 陈传夫
武汉大学信息资源研究中心主任 马费成

2006年10月8日

前 言

数字图书馆至 1994 年由美国国家科学基金会 (NSF)、美国国防部高级计划署 (DARPA)、美国国家宇航局 (NASA) 联合发起的“数字图书馆创始工程”(Digital Library Initiative, 简称数字图书馆 I) 提出以来, 经过了十余年的发展, 该术语已为世界各国所接受, 数字图书馆的研究也日益广泛和深入。2003 年 6 月, 由美国国家科学基金会 (NSF) 主办的 NSF 数字图书馆未来的研讨会 (NSF Post Digital Libraries Futures Workshop), 主题是“无所不在的知识环境”。来自美国各著名大学和研究机构的 40 多名专家学者提交会议论文 40 多篇, 涉及了数字图书馆建设的各个方面。匹兹堡大学信息科学学院的一些教授在向会议提交的一份报告 (Ubiquitous Knowledge Environments: The Cyberinfrastructure Information Ether) 中指出, 数字图书馆的研究重点已经转为 Ubiquitous Knowledge Environment (UKE: 无所不在的知识环境), 或者说数字图书馆将以 UKE 的形式而存在。而 UKE 的存在目的就是大会的副标题: (to provide) Cyberinfrastructure Information Ether (提供计算机基础构造的信息大气)。报告还指出, 真正的挑战来自于如何建立这样一种系统, 它既能支持学术要求和交流, 并且又能高效方便地适应新性能与生产量, 使其运行几乎是不被人察觉的。事实上, 衡量成功与否的最严格的标准, 就是看此构造的设施是否能够以一个十分简单的方式出现在一般用户和高级用户面前。这也就是所谓的“无所不在的知识环境”, 或“信息大气”。同时, 报告对未来数字图书馆的研究也提出了需要关注的一些问题。研讨会指出要建立一个多语言、多媒体、移动的、语义的数字图书馆知识网来检索人类知识, 服务范围从信息服务转向知识服务, 要以前所未有的规模和速度来推动知识进步。数字

图书馆是新世纪发展教育、科研、工程等国家战略的重要组成部分,是评价一个国家信息基础设施水平的重要标志之一。它也是一个庞大的系统工程,涉及理论研究、技术基础、政策法规、人文文化等方方面面。展望未来,数字图书馆将会对社会对人类的精神文明、物资文明起到不可估量的作用。

数字图书馆随着计算机技术、现代通信技术、高密度存储技术的迅速发展而产生,随着研究的不断深入,日益呈现一些相关问题。第一,异构系统之间的互操作问题;第二,数字图书馆在信息表达和检索方面无可读语义信息问题;第三,数字图书馆多以传统的关键词检索为主,其结果往往是与字面意义或某层意义的匹配,越来越无法满足人们对知识的渴望的问题等。为了解决这些问题,我们将本体技术引入到数字图书馆,并且,应用其技术研究开发了基于本体的数字图书馆检索模型。本书充分反映了该模型的全貌,它以历史领域(国共两党第二次合作时期)为突破口,开发了一个实现本体基本功能(本体建库、本体检索、本体推理、本体可视化)的系统,提供了通用程序接口,建立了一个可操作的本体平台;解决了对人物、事件等复杂关系的揭示与检索问题;模型可视化描述形式丰富多样、形象生动,静态和动态的可视化方式使用户操作更加直观、便捷;模型建立了规模较大的本体库,通过推理机制丰富了本体库的内涵,完善了知识发现、组织、管理功能,提高了知识检索深度。专家们一致认为,“研究成果系统、完整,总体上处于国际先进水平,特别是在中文本体领域多种检索模型、推理机制和本体应用结合等方面具有创新性,达到国际领先水平”。

研究成果充分反映本体技术在数字图书馆工程中成功的应用,随着数字图书馆对知识管理的引入,本体提供了一种全新的知识管理模式,较好地解决了知识组织、知识检索等问题。但是目前基于本体的知识管理仅仅局限于静态知识和绝对知识,对于知识的真实性、权威性这些问题也还没有明确的解决方案。为了解决这些问题,作者提出了本体分子理论,本书对该理论从定义到形式化语言描述进行了全面和科学的论证。根据这个理论,作者从解决动态知识、相对知识和多粒度知识入手,构造本体分子知识组织、动态知识的管理和

控制,以及基于本体分子的动态知识的演化过程的可视化描述。研究成果证明了本体分子理论在实践中的可行性、可操作性和科学性。

本书共分六章,第1章,数字图书馆概述,从数字图书馆的定义、特点、作用和机制的范式演化分析了数字图书馆的内涵和本质;第2章,数字图书馆技术,从元数据管理与应用、数字化信息组织结构、数字对象的标识和服务机制等,揭示了实现数字图书馆工程中所应用的主要技术;第3章,本体概论,从本体的定义、语义网的结构、本体的描述语言和本体的逻辑基础分析了本体的哲学定义、各领域的应用以及它的理论基础与基本实现,展现了本体应用前景;第4章,本体技术,从本体库的建设和维护、本体的检索策略、推理机制和处理引擎,分析了本体应用所使用的关键技术;第5章,本体分子理论与应用,作者通过多年的实践和应用本体技术解决问题时遇到的困难,提出了解决困难的途径和方法,即本体分子理论,本章对本体分子理论进行了详细解析和应用实践,揭示了该理论的可行性、可操作性和科学性;第6章,基于本体的数字图书馆检索模型,通过对模型的介绍,描述了本体技术在数字图书馆应用的五大关键技术的细节和步骤,即体系结构解析、语义信息的提取、历史领域资源本体构建、历史领域知识推理机制、基于知识的检索机制等。

本书是国家985二期项目武汉大学信息资源研究创新基地研究项目和国家自然科学基金资助项目(批准号:70373047,70773087)的成果之一,为此,十分感谢武汉大学信息管理学院院长、博士生导师陈传夫教授,十分感谢武汉大学信息资源研究中心主任、博士生导师马费成教授,武汉大学信息资源研究中心副主任、博士生导师胡昌平教授,十分感谢武汉大学信息管理学院副院长、博士生导师李纲教授,副院长、博士生导师方卿教授,副院长、博士生导师王新才教授,十分感谢武汉大学信息管理学院国家985攻关小组领导和武汉大学出版社严红编审对本书的关心、帮助和支持,感谢杨华副编审为本书付出的辛勤劳动。本书的出版是作者多年来研究数字图书馆和本体论及其应用的总结,是我们团队几届博士研究生和硕士研究生共同刻苦研究的结晶。姜赢、孙雨生、张自然、王菲、余传明、杨宁、徐国虎、张继东、丁波涛、沈阳、曾杰、翁丹丹、杜文华、高巾、周义刚、俞思伟、苏

志军、郭立帆、陈琮、陆伟忠、王国育、陈亮、敖青、夏阳、张华、彭翠屏、赵霞、田欣、段巍、张中宁、郭玺鸣、李江宜、张荣辉等,为本书的撰写做出了贡献,在此,表示感谢。

限于水平和时间,加上新技术的不断涌现,书中难免存在错误或不妥之处,恳请广大读者批评指正。

董 慧

2008年1月于武昌珞珈山

目 录

1 数字图书馆概述	1
1.1 数字图书馆的定义	1
1.1.1 国际上数字图书馆的定义	2
1.1.2 我国对数字图书馆的定义	4
1.1.3 数字图书馆与相关概念间的关系	6
1.1.4 数字图书馆与搜索引擎	10
1.2 数字图书馆的特征	14
1.2.1 数字图书馆的平台特征	14
1.2.2 数字图书馆的服务特征	17
1.2.3 数字图书馆的信息组织特征	18
1.3 数字图书馆的作用	21
1.4 数字图书馆机制的范式演化	23
1.4.1 第一代: 基于数字化资源的数字图书馆 (Resource-based Digital Library)	24
1.4.2 第二代: 基于集成信息服务的数字图书馆 (Service-based Digital Library)	25
1.4.3 第三代: 基于用户活动的数字图书馆 (Work-based Digital Library)	26
1.4.4 未来的数字图书馆: 无所不在的知识环境 (UKE: Ubiquitous Knowledge Environment)	28
2 数字图书馆技术	32
2.1 概述	32

2.1.1	数字图书馆技术背景	32
2.1.2	数字图书馆关键技术	35
2.2	数字图书馆技术发展	41
2.2.1	美国数字图书馆技术研究进展	42
2.2.2	英国数字图书馆关键技术研究进展	49
2.2.3	法国数字图书馆关键技术研究进展	51
2.2.4	日本的数字图书馆关键技术研究进展	55
2.2.5	中国数字图书馆关键技术研究进展	57
2.3	元数据管理与应用	66
2.3.1	元数据简介	66
2.3.2	元数据管理	73
2.3.3	元数据应用	76
2.4	数字化信息组织结构与技术	78
2.4.1	数字化信息概述	78
2.4.2	数字化信息组织机制	80
2.4.3	基于 XML 技术的信息内容与结构组织	81
2.4.4	元数据组织	85
2.4.5	知识组织体系的组织	87
2.4.6	信息资源系统的组织	90
2.4.7	信息链接与动态重组	94
2.4.8	信息交流体系的组织	97
2.5	数字对象的唯一标识符技术	98
2.5.1	唯一标识符在数字信息环境的作用	98
2.5.2	唯一标识符的系统框架和性能要求	98
2.5.3	唯一标识符范例	100
2.5.4	唯一标识符开发与应用的趋势	103
2.6	基于信息网格的数字图书馆	104
2.6.1	网格技术的发展	104
2.6.2	基于网格应用的数字图书馆	108
2.6.3	面向服务的数字图书馆信息网格	111

3 本体概述	121
3.1 本体的定义	121
3.2 本体与语义 Web	123
3.3 本体的描述语言标准	127
3.3.1 RDF	127
3.3.2 RDFS	138
3.3.3 OWL	146
3.4 本体的逻辑基础	152
3.4.1 逻辑、推理与知识社会	152
3.4.2 现代逻辑	154
3.4.3 描述逻辑	157
4 本体技术	162
4.1 本体建库	162
4.1.1 本体建库方法论	162
4.1.2 本体建库工具	168
4.1.3 本体半自动建库	176
4.2 本体检索	179
4.2.1 SPARQL 查询语言	179
4.2.2 RDQL 查询语言	186
4.3 本体推理	189
4.3.1 本体和描述逻辑的关系	190
4.3.2 本体推理工具及其应用	193
4.3.3 针对领域的蕴涵知识发现推理	200
4.3.4 本体推理的作用	201
4.4 本体处理引擎	203
4.4.1 Jena API	203
4.4.2 Protégé-OWL API	213
5 本体分子理论与应用	222
5.1 引言	222