

特色数据库建设

理论与实践

The Theory and Practice of Special
Database Construction

© 刘葵波 著



机械工业出版社
CHINA MACHINE PRESS

本书获“集美大学出版基金”资助

中国高等教育文献保障系统（CALIS）三期专题特色数据库项目支持

特色数据库建设理论与实践

刘葵波 著



机械工业出版社
CHINA MACHINE PRESS

本书在广泛调研和考察国内外特色数据库建设与研究现状的基础上,全面阐述了特色数据库的概念与内涵、类型与特征、建设意义与作用;深入系统地研究了特色数据库建设、管理与服务的一系列问题,包括建设原则、基本流程、选题论证、结构体系设计、资源收集与筛选、数据加工与转换、资源标引、检索系统设计与内容发布、技术方案与系统平台、维护与更新、服务内容、宣传与推广、组织与管理等;探讨了特色数据库的版权问题并提出多元化版权解决策略与方案;分析了特色数据库共建共享的必要性、可行性及成功案例,进而系统地论述了共建共享的模式与实施措施;针对特色数据库建设存在的质量问题,提出了特色数据库建设质量控制措施,并探讨了特色数据库的评估办法及指标体系的构建。

本书注重理论与实践相结合,案例分析贯穿在整个研究过程中。通过对特色数据库建设的理论与实践进行系统的研究,以期为我国特色数据库的进一步开发建设与良性发展提供理论参考和实践指导。

本书是教育部“211工程”高等教育文献保障系统(CALIS)三期专题特色数据库立项项目“水产科技数据库”、“陈嘉庚研究数据库”和福建省教育厅科技项目“水产科技数据库”的研究成果。

图书在版编目(CIP)数据

特色数据库建设理论与实践/刘葵波著. —北京:
机械工业出版社, 2011. 12

ISBN 978-7-111-37357-5

I. ①特… II. ①刘… III. ①数据库系统 IV. ①TP311.13

中国版本图书馆CIP数据核字(2012)第018132号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

策划编辑 伏开玲 责任编辑 罗子超

北京兴华昌盛印刷有限公司印刷

2011年12月第1版·第1次印刷

140mm×203mm·11.875印张·282千字

标准书号: ISBN 978-7-111-37357-5

定价: 30.00元

序

在当今数字化、网络化时代，作为收藏信息资源的图书馆，其资源建设与服务应与时俱进，不断发展，才能满足用户不断扩大且多样化的信息需求。特色数据库是传统图书馆的特色馆藏数字时代的另一种表现形式，是图书馆数字信息资源建设的核心内容与发展方向，是开展网络信息服务的主要资源。特色数据库的建设有利于充分发挥图书馆的特色资源优势，有效弥补商业数据库的不足，最大限度地满足了不同用户的信息需求，在更广的范围内实现资源共享，在更高的层次上服务于用户。同时，还有利于树立图书馆品牌地位，增强图书馆在社会中的竞争力。因此，近年来，许多图书馆特别是高校图书馆围绕着特色馆藏、优势学科、地域资源开展了卓有成效的特色数据库建设，开发出了题材丰富、形式多样的特色资源数据库。

在特色数据库建设与实践活动广泛开展的同时，相关的理论研究也得到了迅速发展。业界不少学者都从不同角度对特色数据库的建设进行过不同程度的研究，也取得了一定的成绩。但目前的研究还显得比较零散，缺乏系统性和完整性，而且对特色数据库建设的核心和重点问题缺乏进一步深入的研究。作者正是在这一背景下撰写了本书。该书的出版填补了这一领域

研究的空白，是一部具有创新意义的学术专著。

该书对特色数据库建设的理论与实践进行了全面、系统和深入的研究。作者从大量调研着手，在广泛考察国内外特色数据库建设现状的基础上，详细而客观地分析了各系统、各地区特色数据库建设的特点及存在问题；首次深入地探讨了特色数据库的概念、内涵、特征等基本理论；从建设原则、基本流程、选题与架构、资源的收集、加工、标引、检索、发布、技术方案与系统平台、维护与更新、服务与推广、组织与管理等方面系统全面地论述了特色数据库建设、管理与服务的一系列问题；在我国现行的法律框架内，对特色数据库建设与使用过程中涉及的版权问题进行了较为透彻的分析，进而根据相关法规的适用性，有针对性地提出了多元化的版权解决策略与方案；对特色数据库的共建共享问题进行了深入的思考，并结合案例对共建共享的实现机制与模式提出了可行性思路 and 措施；在对特色数据库建设存在的质量问题进行深刻剖析的基础上，积极探讨了特色数据库建设的质量控制措施及特色数据库的评估。

该书作者积极致力于馆藏特色资源的开发与建设，其所在的集美大学图书馆已自主开发建设了多个数据库，其中“水产科技数据库”和“陈嘉庚研究数据库”都是福建省内比较优秀和有影响力的特色数据库，并且还是 CALIS 特色数据库建设项目和福建省教育厅科技项目。作者在特色数据库建设与管理方面具有丰富的实践经验，并在实践过程中掌握了大量第一手数据与资料，为其开展的研究奠定了坚实的基础。

该书选题新颖，结构层次分明，逻辑严密，研究方法得当，理论探讨和应用研究相结合，研究和分析结果具有较强的创新

性、说服力和可操作性。其学术价值在于反映了我国目前在这一研究领域的新成果、新观点、新看法，拓展和丰富了数字图书馆的理论体系，对业界的理论研究提供了有益的参考借鉴作用。其应用价值在于为我国特色数据库的进一步建设与发展提供了良好的理论参考和实践指导。可以说，该书熔学术性与应用性于一炉，是研究特色数据库建设理论与实践不可多得的优秀专著。我相信，该书的问世必将有利于促进我国特色数据库事业形成更加完善的建设体系、服务体系和共享体系。

福州大学信息化建设办公室主任
福州大学信息管理研究所所长
教授、研究馆员、博士生导师
张文德

目 录

序

第一章 绪论	1
第一节 研究背景与意义	1
第二节 研究内容与方法	4
第二章 特色数据库概述	7
第一节 特色数据库的概念	7
第二节 特色数据库的特征	18
第三节 特色数据库的类型	21
第四节 特色数据库建设的意义与作用	42
第三章 特色数据库的国内外建设现状	55
第一节 内地图书馆特色数字资源建设现状	55
第二节 台港澳地区图书馆特色数字资源建设现状 ...	101
第三节 国外图书馆特色数字资源建设现状	108

第四章 特色数据库的建设	117
第一节 特色数据库建设的原则与基本流程	117
第二节 特色数据库建设的前期论证与选题	123
第三节 特色数据库的结构体系设计	129
第四节 特色数据库资源的收集与筛选	136
第五节 特色数据库的数据加工与转换	145
第六节 特色数据库的资源标引	160
第七节 特色数据库检索系统的设计与内容发布	189
第五章 特色数据库建设技术方案与系统平台	204
第一节 特色数据库建设的相关技术	204
第二节 特色数据库的技术方案	209
第三节 特色数据库的系统平台	213
第四节 特色数据库的硬件平台	238
第六章 特色数据库的服务与管理	240
第一节 特色数据库的服务	240
第二节 特色数据库的维护与更新	246
第三节 特色数据库的宣传与推广	251
第四节 特色数据库建设的组织与管理	258
第七章 特色数据库的版权问题	272
第一节 特色数据库版权问题现状	272
第二节 特色数据库建设与使用过程中涉及的版权问题 分析	273

第三节	特色数据库多元化版权解决策略与方案	281
第四节	积极参与和影响著作权立法进程	290
第八章	特色数据库的共建共享	293
第一节	特色数据库共建共享的必要性和可行性	293
第二节	特色数据库共建共享的成功案例	296
第三节	特色数据库共建共享的模式	299
第四节	特色数据库共建共享的采取措施	302
第九章	特色数据库的质量控制与评估	311
第一节	特色数据库建设存在的质量问题	311
第二节	特色数据库建设的质量控制	312
第三节	特色数据库的评估	315
附录	部分特色数据库网址一览表	334
参考文献		363

第一章 绪 论

第一节 研究背景与意义

特色数据库是 20 世纪 90 年代中后期,在全球信息技术革命和数字图书馆建设热潮的背景下出现的新概念。它是指图书馆或相关信息机构依托馆藏特色文献信息资源,针对特定用户的信息需求,对某一学科或某一专题的资源(包括分散的、传统的馆藏文献资源,异构的数字馆藏资源以及互联网上的相关信息资源)进行收集、分析和筛选,再按照一定的标准和规范将其进行加工处理、标引、组织和存储,通过局域网或广域网进行传输和发布,提供检索和全文传递服务,以满足用户个性化需求的文献信息资源库。

特色数据库首先必须是图书情报机构自建的资源库,其次是具有明显特色。只有同时具备“自建”与“特色性”这两个要素的数据库,才可称为特色数据库。特色数据库与特色馆藏、自建数据库之间的关系是:特色馆藏包含但又不仅限于特色数据库,特色数据库是特色馆藏的重要组成部分,是体现馆藏特色的重要资源;自建数据库包含但又不仅限于特色数据库,特色数据库是自建数据库的重要组成部分;自建数据库不完全都属于特色馆藏的范畴,只有其中的特色数据库才基本属于特色

馆藏的范畴；特色数据库基本都属于自建数据库。

国外图书馆特色数字资源建设起步较早，虽然国外并没有“特色数据库”这一名称和提法，但从 20 世纪 90 年代起，美国国会图书馆和斯坦福大学、卡内基—梅隆大学、加州大学伯克利分校、依利诺伊大学、密西根大学、加州大学圣芭芭拉分校等名校图书馆开始了特色数字资源的建设，许多知名图书馆都建有自己的特藏数据库。在美国之后，英国、法国、德国、加拿大、意大利、荷兰、日本等各个国家也都启动了图书馆资源数字化建设工作。国外大学图书馆的特色资源建设都较为成熟，有些大学图书馆更是以其特色馆藏而闻名于世，他们通常都建立了特色馆藏收藏部门，并进行数字化处理和管理，为本校的师生、本州的乃至全世界的用户提供查阅服务。

在全球数字化发展的趋势下，我国台湾地区高度重视数字化建设。2000 年和 2002 年，台湾地区“行政院”分别启动了“国家典藏数位化计划”及“数位典藏国家型科技计划”。在这一项目的推动下，台湾地区图书馆的特色数字资源建设取得了丰硕的成果，特别是台湾“国家图书馆”（台湾广域数位图书馆）和台湾大学图书馆的特色数字资源建设成绩斐然。港澳地区的图书馆为促进资源的利用与共享，也开展了不同程度的馆藏特色资源数字化工作或建立了特藏书目数据库，比较突出的有香港中文大学图书馆、香港大学图书馆和澳门中央图书馆等。

大陆地区图书馆特色数据库的建设起步于 20 世纪 90 年代的“中国国家试验型数字式图书馆”试验项目，这一项目拉开了我国建设特色数据库的帷幕。20 世纪 90 年代中后期，在数字图书馆建设热潮的影响下，众多的图书情报机构，特别是高校图书馆都充分认识到特色数据库在本馆文献资源建设和数字图书馆建设中的重要意义，在科学规范引进电子资源的同时，纷纷开展了特色数据库的建设。经过多年的探索和发展，我国各系统、



各类型图书馆的特色数据库的建设已取得阶段性的成果,有的已经取得了非常好的效果,并有继续规模发展的趋势,可谓是方兴未艾。从建成和在建的专题特色数据库来看,高校图书馆、公共图书馆以及科研院所图书馆这三大类型图书馆的特色数据库建设呈现不同的特点,可谓各有千秋、各具优势。其中,高校图书馆系统是我国特色数据库开发和建设的主力军。特别是在中国高等教育文献保障系统(CALIS)“九五”专题特色数据库一期、“十五”专题特色数据库二期及2011年2月启动的三期项目的有力推动下,高校特色数据库建设更加蓬勃地发展,已初步建成一定规模的,具有中国特色、地方特色、高等教育特色等资源特色,而且方便实用、技术先进的专题特色数据库群,这些专题特色数据库面向全国高校提供多层次和个性化的特色资源服务。

数字化和特色化是当代图书馆馆藏建设的两大主流,而特色数据库建设正好是这两大主流的自然融合。因此,自建专题特色数据库是图书馆数字资源建设的重要内容和发展方向,是数字图书馆建设的重要组成部分,对数字时代图书馆的发展与建设具有积极的意义和重要的作用。

特色数据库建设是保存和开发利用文献资源的基础工作,有利于充分开发和保护特色馆藏资源,并提高其利用率和使用价值;特色数据库是满足信息时代用户多样化和个性化需求的有效手段,并能够优化和发展特色馆藏,深化和拓展特色服务;特色数据库能为地方经济和文化建设提供信息支撑服务,能为高校学科建设和教学科研提供重要资源保障;特色数据库是信息时代图书馆自身生存与发展的需要,更是文献信息资源共建共享的必然要求。

随着特色数据库建设与实践活动的广泛开展,对特色数据库的理论研究也不断升温,特别是近年来的研究更加活跃。国

内外有关专家学者也对特色数据库的研究给予了一定的关注。目前,各方面有价值的研究成果已有不少,这些成果为人们进一步的研究和探讨提供了有益的借鉴和参考。但现有的研究成果大多集中在论述特定的特色数据库或特色数据库某方面的问题,包括对某一专题特色数据库建设的个案研究;对某领域、某区域范围特色数据库建设与管理的研究;对特色数据库建设的必要性、建设原则、建库技术、版权、质量控制、建设标准、评价等方面的研究等。总体而言,目前业界对特色数据库建设理论与实践的研究还不够全面、系统和深入;对特色数据库的基本概念、内涵及特征至今尚未完全阐明,还需要进一步厘清;对特色数据库的结构体系设计、资源收集、资源标引与检索、共建共享机制与模式、建设效益、版权障碍、可持续发展等突出问题和难点问题的分析研究,以及对这些问题提出相应的对策还相对不够,有待继续深入与拓展。

正是在这种背景下,本书力图在广泛调研和考察国内外特色数据库建设实践与研究现状的基础上,从理论与实践两个层面对特色数据库建设进行全面、系统和深入的研究,阐明特色数据库的概念、内涵、类型及特征等基本问题,对特色数据库建设中的一系列突出问题和难点问题进行深入的分析 and 研究,并有针对性地提出解决问题的原则、思路、方案及实施措施,以期为我国特色数据库的进一步开发与建设与良性发展提供理论参考和实践指导,进而促进我国特色数据库建设事业呈现更好的发展和应用前景,形成更加完善的特色数据库建设体系、服务体系 and 共享体系。

第二节 研究内容与方法

本书共九章,主要内容包括特色数据库的概念与内涵、类



型与特征、建设意义与作用；国内外特色数据库的建设现状与特点，并指出了存在的问题；特色数据库建设、管理与服务的一系列问题，如建设原则、基本流程、选题论证、结构体系设计、资源收集与筛选、数据加工与转换、资源标引、检索系统设计与内容发布、技术方案与系统平台、维护与更新、服务内容、宣传与推广、组织与管理等；特色数据库建设与使用过程中涉及的版权问题，并提出多元化版权解决策略与方案；特色数据库共建共享的必要性、可行性及成功案例，进而探讨了适合我国国情的共建共享实现机制、模式与实施措施；针对特色数据库建设存在的质量问题，提出了特色数据库建设的质量控制措施，进而探讨了特色数据库的评估办法及指标体系的构建。

本书理论探讨与实践研究相结合，以图书馆学、情报学、信息学理论与方法为主要研究工具，借鉴系统论、管理学、逻辑学、计算机网络科学等诸多理论与方法，同时，案例分析贯穿在整个研究过程中。

本书的研究主要采用了以下几种方法。

1. 调查研究法

调查研究法主要是采用网络调查法。在本书近三年的写作过程中，对国内外一百多家图书情报单位的两百多个特色数据库网站，及其他数十个相关网站逐个进行了访问，对部分有代表性的特色数据库网站进行了跟踪访问。通过对众多特色数据库网站有计划、有目的、广泛而系统地访问和使用，掌握了大量的素材和案例，并对调查搜集到的资料进行分析、综合、比较、归纳，从而为研究提供了有力的支撑。

此外，还通过对各类型图书馆的实地考察、访谈专家学者和同行、参加专题研讨会等方式获取相关资料，并以此为基础进行分析论证。



2. 经验总结法

本书作者近年来主持了“水产科技数据库”、“陈嘉庚研究数据库”、“信息参考库”等若干个特色数据库的建设,积累了较为丰富的实践和研究经验。通过对建库活动中的各种情况进行归纳、分析与总结,产生了丰富的第一手资料和翔实可靠的数据,并使之系统化和理论化,从而为本书的写作奠定了坚实的基础。

3. 文献研究法

从各种期刊论文、学位论文、电子图书等数据库中以及传统书刊资料中,查找和搜集有关特色数据库和数字图书馆的资料信息,通过分析梳理这些研究成果,全面系统地了解 and 掌握目前国内外图书馆界有关特色数据库研究领域的历史、现状与最新进展,有助于研究的设计和规划,提供研究对照,拓宽研究视野,提高论证水平,并在继承的基础上创新发展。

4. 系统分析法、比较分析法

通过对国内外不同类型、不同地区、不同单位的图书馆特色数据库建设情况进行系统分析和比较分析,探求其异同与内在规律,以便从系统论的角度出发对特色数据库的建设体系、服务体系和管理体系进行分析和研究。

5. 案例分析法

本书对特色数据库的建设情况进行了大量的调查研究,既有对某个特定专题特色数据库的分析与研究,也有对某类型特色数据库的评析,还有对某系统或某地区特色数据库建设情况的调查研究。在对这些有代表性的特定对象进行案例分析的基础上,有针对性地对特色数据库建设中的有关问题和现象进行深入的研究,并提出解决方案。



第二章 特色数据库概述

第一节 特色数据库的概念

一、数据库的概念及类型

随着计算机技术、通信技术和网络技术的飞速发展,图书馆进入了以信息资源共享为目标的数字图书馆阶段。文献信息资源数字化、传递网络化、信息利用共享化、信息实体虚拟化已是数字图书馆的基本特征。数据库技术是现代计算机信息系统和计算机应用系统的基础和核心。在网络化和数字化的环境下,数据库是信息资源的主要组织方式、存在形态和利用形式。信息资源是以数据库方式存储,并以网络服务方式进行传播。数据库建设是现代信息产业的基本建设工程,是整个社会信息资源开发和利用的基础,是现代图书情报信息系统的重要支柱之一。一个国家拥有数据库的数量,能用数据库提供服务的程度,是国家经济实力、文明程度和科技水平的重要标志。

所谓数据库,是在计算机可读介质上,按一定的结构和规则组织起来并存储,借助于检索系统随时为用户提供服务的相关信息的集合。通俗地说,数据库就是以某一特定方式组织和存放数据的仓库。数据库系统通过自动化手段来实现资源的加工、管理、服务和产品生产。它以先进的数据处理技术、深加



工的信息以及与网络的完美结合，成为数字信息资源的主体。数据库的本质具有三个要素：相关信息、共同的存取方式和一定的组织方式、共享。

计算机信息检索系统的数据库，按记录形式可以分为文献型数据库和非文献型数据库两大类，如图 2-1 所示。

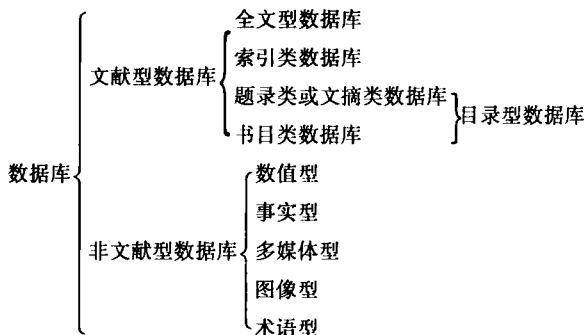


图 2-1 数据库类型示意图

文献型数据库是一种直接提供文献线索或文献原文的数据库，收集的记录是根据书刊、技术报告、学位论文、会议论文、专利、报刊新闻等加工而成的一次或二次文献，可分为目录型数据库和全文型数据库。其中，目录型数据库也就是二次文献数据库，是存储图书目录、论文题录、文摘等书目线索的数据库，又可以分为索引类数据库、题录类数据库或文摘类数据库、书目类数据库等。例如，“世界各国标准索引”（World wide Standards Service Plus, WWP）为索引类数据库；“科学引文索引”（Science Citation Index, SCI）、电子版“全国报刊索引”为题录类数据库；电子版“经济学文献库”（EconLit）为文摘类数据库。全文型数据库也就是一次文献数据库，是存储文献全文或部分内容的数据库。它提供原始的文献数据，集检索功能与浏览原文功能为一体，是目前最便捷、最具吸引力的和最具潜力的数字资源，如中国知网、超星数字图书馆、EBSCO 等。

